University of Reading

Responsibility for Climate Change: Collective Harm, Individual Accountability and Radical Moral Revisionism

PhD in Philosophy

Department of Philosophy

Bennet Francis

December 2020

[Blank Page]

Declaration:

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

BENNET FRANCIS

[Blank Page]

Abstract

Climate change is an instance of the problem of collective impact. This is a problem in normative ethics, which arises when the actions of many individuals together produce some morally significant outcome, but no agent is apparently an appropriate object of accountability for it. Given climate change is a disastrous, agent-caused, foreseen harmful outcome, many people share the judgment that some agent or agents ought to be accountable for that harm, yet no individual's behaviour is apparently of sufficient moral significance for assignments of accountability, and corresponding assignments of remedial duty, to be appropriate. The problem has led some theorists to the drastic conclusion that our existing moral concepts are fundamentally unsuited to the complex global structure of interpersonal relations in which we today find ourselves, and that we therefore need to create radically new moral concepts. The thesis offers a critical taxonomy of responses to the problem in the existing literature, before focusing on three very different types of response: defences of individual direct duties to reduce personal emissions, defences of participatory duties grounded in minimal forms of shared agency, and radical moral revisionism about the content of our individual duties. While the first of these kinds of response fail, the thesis develops and defends a version of the second, based on the idea of quasi-participatory accountability. This is the sense in which individuals can be accountable for the outcomes associated with group behaviours when those individuals identify with those behaviours and regard themselves as participating in collective action, even in circumstances when there is no actual coordination between individual agents. The thesis argues that the radical revisionist approach relies on a deeply flawed understanding of the philosopher's role in influencing moral attitudes. It is therefore fortunate that an approach based on quasi-participatory accountability renders radical revisionism unnecessary.

[Blank Page]

Acknowledgements

I would like to thank my supervisors Brad Hooker and Rob Jubb, who have guided my thought in too many places to innumerate. Thanks also go to Elizabeth Cripps for generous comments on draft chapters and discussion.

I am grateful for scholarship funding from the Leverhulme Trust. I am equally grateful to my fellow Leverhulme Scholars in Climate Justice for invaluable advice, support and friendship. In particular, I would like to thank the politics group consisting of Alex McLaughlin, Joshua Wells, James Draper, Lydia Messling, Livia Luzzatto and Adam Pearce.

I would like to thank the scholarship programme's founding director, Catriona McKinnon, for fostering this community, and for orchestrating the many opportunities that made it so rewarding.

I would also like to thank members of the REAPP group for fruitful discussions of work and related research, especially Bradley Hillier-Smith, Steven Wu and Aart Van Gils. Some of these discussions informed chapters of this thesis, but mostly they were just a lovely way to spend an afternoon.

I am deeply grateful to my parents, John and Helen, not only for their unwavering support and patience, but for helping to keep my mind in balance, and to Katya, for being my constant companion through sweltering evenings in BNF when we would rather have been out dancing.

Finally, I would like to express my deep sadness that my maternal grandfather, Bedrick, and my paternal grandmother, Eileen, passed away while I was working on this project. Bedrick started out as an academic, while Eileen in later life acknowledged her regret that she had not had the opportunity to attend university. I hope I have continued on their path in some small way.

[Blank Page]

Contents

[Blank Page]

1. Introduction

Climate change is caused, at least in part, by millions of agents, widely dispersed across time and space, behaving normally. They break no laws, offend against no common-sense moral principles, violate no conventions. Yet together, they give rise to a transformation of the earth's climate system so profound that the capacity of these systems to support the forms of life humans and other animals enjoy today could be placed in jeopardy. Climate change is already doing significant, demonstrable damage to the wellbeing of people all over the world, and to the integrity of the societies in which they live. And the more the climate is changed, the more people will suffer.

This is a dire problem from the perspective of many different fields of human understanding. It is a technological problem, a problem for all spheres of politics from the most local to the most global, for law, for economic and social theory, and for art. Among the multitude of problems thrown up by climate change is the peculiar one it presents for the field of normative ethics. Whatever else it may be, climate change is a grievous instance of human beings gravely harming other human beings. When we learn that under existing conditions, as many as 700 million people could be displaced by increased water scarcity due to the changing climate before 2030 (Hameeteman 2013 8) and that by 2040 a quarter of the world's children will be living under extremely high water stress (UNICEF 2017 2), when we hear that the state of Kiribati has already purchased land with a view to transplanting their entire nation to territory held by other countries, to preserve their society as best they can as their homeland is destroyed forever,[1] when we watch in horror as our television screens are filled with images of the Australian bushfires of 2019-20, which

[1] See http://www.climate.gov.ki/2014/05/30/kiribati-buys-a-piece-of-fiji/ (retrieved 8 Oct. 2020).

caused, at one estimate, 100 billion Australian dollars of property damage,[2] and killed 451 people along with more than 1 billion animals,[3] and when we know furthermore that *people are doing this to each other*, we are left with the overwhelming impression that something not just tragic, but monstrous is occurring.

There are a number of ways of framing this moral problem, as we shall shortly set out. To begin laying out the territory, we can start with idea that because climate change is a disastrous, foreseeable agent-caused outcome, a serious failure of practical reasoning has occurred, and yet apparently no agent has sufficient reason to act any differently: a paradox. To illustrate, suppose I have to make an everyday decision of a kind that will determine whether some additional quantity of greenhouse gas (hereafter GHG) is emitted. For example, when ordering a taxi, I can choose between ordering a petrol or diesel taxi or ordering a taxi with low emissions. Choosing the greener option has some cost attached to it – perhaps it costs a pound more, or perhaps I have to wait slightly longer. What reason do I have to choose the greener option? Do I have any reason at all? If I fail to choose the greener option, have I neglected some duty? A duty to whom? Am I perhaps rationally required to take the cheaper option? It seems natural to conclude that I have no reason to choose the greener option, and it is even less plausible that I have a decisive reason. The choice might seem trivial, but this is precisely the point: millions of such trivial choices, made by millions of people all over the world, together cause serious harm through climate change.

[2] See https://edition.cnn.com/2020/01/10/perspectives/australia-fires-cost/index.html (retrieved 8 Oct. 2020)

[3] 34 people were directly killed by fire, and 417 by smoke inhalation. See (Borchers Arriagada 2020); see https://www.usgs.gov/center-news/geoscience-australia-s-oliver-discusses-use-landsat-during-country-s-historic-fires?qt-news_science_products=1#qt-news_science_products (retrieved retrieved 8 Oct. 2020)

Some people conclude that climate change has laid bare 'the impoverishment of our system of practical reason' (Jamieson 2014 8). What it reveals, these philosophers say, is that our existing moral concepts have failed. Many philosophers, as well as scholars in the environmental science literature (see Singer, e.g. 2005, Jamieson 2014, Markowitz & Shariff 2012) argue that because our moral framework emerged during a period of human history when we lived in small groups, common-sense morality was not meant for understanding the new and complex ways in which people can negatively affect the lives of others in the contemporary world. For this reason, we are simply not "hard-wired" to respond adequately to climate change.

This claim has a purely psychological reading, and a specifically moral reading. On the one hand, as Dale Jamieson puts it, 'evolution built us to respond to rapid movements of middle-sized objects, not the slow build-up of insensible gases in the atmosphere' (Jamieson 2014 4). The thought here is that we find it difficult to become emotionally exercised by climate change because it is invisible in our day to day lives. The summers get hotter and the winters wetter year on year, and after a time we start to forget things were ever different. This makes it more difficult to feel the emotional pull of the call to action.

It is not just the invisible and novel character of climate change that is supposed to make it a particular problem for normative ethics, however. There are plenty of forms of wrongdoing which our psychological development as a species could not possibly have "hard-wired" us to detect, yet we have no problem condemning when it is pointed out to us. Think of violations of data privacy, for instance, or inflating share prices by submitting false accounts. Our distant ancestors would not have been able to identify these wrongs, but they would likely have had no difficulty recognising them as wrongs given the proper explanation of the unfamiliar concepts involved. The more troubling argument is that although it strikes us that wrongdoing is involved in anthropogenic climate change, we are

unable to identify any wrong that has taken place, not because we do not know how to apply our moral concepts to this unfamiliar territory, but because the right moral concepts do not exist. The evolution of our moral psychology has not filled our conceptual toolbox with the right equipment. We are, on this view, at 'the Frontiers of Ethics' (Jamieson 2014 144), and it is up to us to find, and indeed create, new paths.

This thesis argues that we need not be so pessimistic. Both our moral concepts, and our moral emotions are suitably well-fitted to respond to the problem of climate change, once we have the right understanding of the relationships in which each of us is engaged. Individual contributors to GHG emissions both can and should regard themselves as bearing responsibility for climate change. Each individual should regard the harms caused by specific carbon-intensive economic structures in which they are involved as *their mess*.

Framings of the Problem

The No-Difference Problem

One of the most influential framings of this problem has been the problem of collective harm, sometimes called the problem of inconsequentialism, or the no-difference problem. This is an abstract problem that has been perennial in moral philosophy, at least since Jonathan Glover and M.J. Scott-Taggart's characterisation of the "no difference argument" in the 1970s (Glover and Scott-Taggart 1975), and in political and economic theory, at least since Anthony Downs's description of the voting paradox two decades earlier (Downs 1957). As it is a matter of contention whether climate change itself counts as an instance of the problem viewed under this framing, let us consider first a somewhat more artificial case, adapted from (Kagan 2011). Suppose a factory produces some pollutant that is damaging to health in sufficiently large quantities, but harmless in small quantities (to make the case clearer still, we can even suppose it is beneficial in small quantities, as small concentrations of fluoride ions added to drinking water can be good for the teeth). One factory produces

a certain amount of this pollutant, which, in a predictable fashion, ascends into the atmosphere and disperses evenly all over the globe, before descending back down to a level where it can be inhaled. Because it is evenly dispersed, no one person inhales more than a single molecule, which has no effect whatever on their health. But thousands of factories all over the world are releasing the pollutant in similar quantities to the first factory, and therefore in combination are causing a significant burden of ill-health to people all over the world. Each factory owner can claim that she does nothing wrong, because she makes no difference: were she to refrain from polluting, the global burden of ill health as a result of the pollutant would be exactly as bad as it would be with her pollution. Nevertheless, it looks like the factory does wrong by polluting. This is a potentially paradoxical result.

We are faced with three questions: i) is it true that the individual makes no difference in cases such as this? ii) Does it follow from the fact that the individual makes no difference that no wrongdoing has occurred, or can wrongdoing be established in other ways? iii) Is climate change a case of this form? Kagan, who presented a version of the case, argued that it is false that the individual factory owner makes no difference, because the impact of one molecule of pollutant on the lungs of one individual can still be considered harm, though minute, and that many such minute harms add up to serious harm (a response that follows Parfit 1984 80). As Julia Nefsky points out, however, the individual factory owner has the same impact whatever the other factories do (Nefsky 2011 372). Imagine a second case, one in which no other factories pollute. The molecules produced by the single factory will be dispersed among millions of people, they will make no difference to anyone's health - indeed they will produce benefits, on our original supposition - and no harm will occur. Under these circumstances, it cannot be reasonable to attribute wrongdoing to the factory owner. But if, in the initial case, the wrong the factory owner does is determined by the imperceptible amount of harm the pollution she emits does to each of the people who

inhale it, multiplied by the number of people, this figure is the same whether or not the other factories also pollute. If she makes no difference in the second case, it is difficult to see why she should be regarded as making a difference in the first.

Kagan introduces the no-difference problem as a problem for consequentialism in particular, but as Nefsky argues, non-consequentialists also have good reasons to be troubled by it. Though, as this thesis will argue, the claim that no difference-making implies no wrong-doing is indeed false, this result is not simply a function of the denial of consequentialism. As Chapter 2 will lay out, a number of standard non-consequentialist approaches are also unable to resolve the problem. On the question of the relationship between this artificial framing of the problem and climate change itself, it should be noted that even a philosopher who allowed that there were true no-difference cases, and even one who accepted that Kagan's factory case was one of them, may nevertheless deny that climate change was a true no-difference case. As will be shown in Chapter 3, it is reasonable to conclude that an individual act of producing GHG emissions makes a difference in terms of harm, because it causes expected harm. In other words, it increases the risk of harm. This realisation does little to solve the problem however, as each contributor to climate change can still argue they do no wrong. This is because, in all but the most wanton instances of producing GHG emissions, each does what she has most reason to do. The small expectation of harm the individual produces is easily outweighed by the benefit she expects to derive from the carbon-intensive activity. In our taxi case, for instance, the net expected value of choosing the cheaper rather than the greener taxi is likely positive.

There is another route to the conclusion that individual mitigation efforts make no difference, one that appeals to the way the decision to reduce one's emissions interacts with the market behaviour of other agents. A great many of decisions standardly viewed as contributing to an individual's 'carbon footprint' do not directly cause emissions to be

produced. The most that can be said is that they contribute to demand for carbon-intensive services. When I turn on an electric light, I do not cause additional fossil fuels to be burned. The electricity company has already determined the load that will be put through the grid based on the average demand for the time of year and time of day. If I make a habit of reducing energy usage, and thousands of others in my area do the same, the combination of all these choices might cause the energy companies to change their estimates for demand, meaning less fossil fuel might be burned in future. But my act of turning off the light does not change the amount of energy already in the grid, and therefore makes literally no difference in itself.

Market behaviour may also mean my attempts at mitigation have negative feedbacks. A complaint raised against the Kyoto Protocol, which classified countries into Annex 1 parties, which had binding emissions-reductions targets, and non-Annex 1 parties, which did not, was that the imposition of emissions reductions targets on some countries but not others would give unregulated countries a competitive advantage, meaning carbon-intensive industries would simply move from, say, the USA, to, say India or China. This would not only be detrimental to the economy of the USA, but would may also have a negative impact on climate change mitigation, as the lower regulatory standards in these countries would allow industries to operate even more inefficiently, or so it was claimed (see for example Murkowski 2000, Coon 2001).

A similar phenomenon might arguably be produced by the voluntary decision of various individuals to reduce their emissions. Suppose, for example, that enough ecologically minded travellers decide to take the Eurostar train between London and Paris instead of flying that one or more airlines decide that it is no longer profitable to lay on so many flights. The knock-on effect would be that air traffic slots were freed up, which might be allocated to more environmentally damaging long-haul routes. This would mean, not only

would an individual traveller's decision to choose the greener option not make a positive difference, it could actually make a negative difference, or at least be part of the cause of net-negative impact. These considerations should already arouse the suspicion that if there is wrongdoing involved in participating in the fossil fuel economy, it is not a result of individual difference-making.

A note of terminological clarification: the problem of collective harm is clearly related to the so-called problem of collective action, but it is distinct from it. The problem of collective action is a problem for rational choice theory. Such problems occur in situations in which the most rational choice for all is not necessarily the most rational choice for each. A classic example is the regulation of so-called "common pool resources", resources that have a renewable "flow" component and a depletable "stock" component. If all agree to consume only the flow component, the resource can be used indefinitely, but if the stock component is depleted the resource may be lost forever. It is collectively rational to ensure only the flow resources are consumed, but it is individually rational for each to consume to a level that, were everyone to consume to that level, the stock would be depleted. This is a problem for rational choice theory because, by doing what is individually rational, each individual undermines their own interests – a paradoxical result. If others are likely to co-operate, then self-interest is better served by "free-riding" on their compliance, gaining more resource at no additional cost, but if everyone, reasoning similarly, attempts to "free-ride", the individual's self-interest is no longer served. Unlike the problem of collective harm, the problem of collective action is not a problem in normative ethics. It may be, for instance, that there are considerations of fairness that militate against the over-consumption of resources, but this observation is irrelevant from the perspective of rational choice theory, which is concerned with modelling and predicting patterns of behaviour on the assumption that actors aim to satisfy a maximum number of their preferences.

The Responsibility Gap and the Collective Duty Gap

It is at least a controversial question whether climate change has the structure of a true no-difference case. It perhaps therefore more illuminating to view the problem under a different framing, as a *responsibility gap*, or as a *collective duty gap*. The thought here is that because climate change is human-caused, avoidable, and seriously harmful, there is a sense in which some agent or agents ought to be responsible for the harm caused, yet every agent involved in producing the harm has a reasonable excuse absolving them from accountability. This again can be viewed as generating a paradoxical conclusion, acutely put by Dale Jamieson in the following terms: '[t]oday we face the possibility that the global environment may be destroyed, yet no one will be responsible' (Jamieson 1992).

The type of responsibility we are interested in here is what Tony Honoré (1999) calls outcome responsibility. This is the sense in which a particular agent can be 'credited or debited' with a particular outcome. Jamieson's worry, in other words is that there is that there may be no one to whom the outcome can be attributed and from whom we can seek redress, or no one of whom we can say that they failed to meet the required standard of behaviour. Whether someone can be credited or debited with an outcome is an input to the determination of where benefits and burdens justly fall. Thus when we say that some agent or agents 'ought' to be outcome-responsible for climate change, what we are saying in effect is that it is unfair that the burdens of climate change should fall upon its victims, because there are, or should be, other agents to whom these burdens should more appropriately be assigned. As David Miller writes, our interest in outcome responsibility arises in part from our desire to 'protect people from the side-effects, intended or unintended, of other peoples' actions' (Miller 2007 89).

For an entity to be assigned outcome responsibility for some impact, it must be a result of their agency to some extent. Outcome responsibility, then, is related to causal

responsibility, but also comes apart from it. Causal responsibility is employed in the explanation of why certain events occurred. It can thus be assigned both to agent-causes and to non-agent causes. When a magistrate has to determine whether a car accident was caused by reckless driving, or by ice on the road, she is making a determination of causal responsibility. If she determines that it was caused by reckless driving, she is making a further determination of outcome responsibility, as if someone's reckless driving was the cause of an accident, it is fairer to reassign the costs of the accident to that person, than it is to allow the costs to fall where they lay, on the accident's victims. Outcome responsibility does not require the agent to be involved in the physical production of an event; it can be applied in cases where an event occurs because an agent fell below the socially required standards of behaviour for someone in their position. For example, if a patient dies from an easily treatable disease while in a doctor's care, the doctor may be outcome-responsible for the patient's death, though the physical causes of the patient's death do not involve the doctor.

Some assignments of outcome responsibility involve moral guilt, while some do not. If a road accident is the result of my intentional reckless driving, I am blameworthy. If an accident is the result of my driving recklessly because I am afraid for the life of my injured spouse whom I am rushing to hospital, then I am outcome responsible, but not blameworthy. This is reflected in the law by the fact that I could not be criminally convicted for causing the accident; nevertheless I, or my insurance company, would have to pay for the damages.

The puzzle of collective harm viewed as an outcome responsibility gap can be framed in terms not of difference-making but of control over contribution.[4] It is unfair, the argument

[4] In this view I take my cue to a large extent from Robert Jubb (see Jubb 2012), whose work is in turn heavily influenced by David Miller (2007 93) and Christopher Kutz (2000). Jubb considers the conditions for *liability* for contribution to a collective harm, which is related to but distinct from

would go, that burdens should be imposed on people who were powerless to avoid having them imposed upon them. As Robert Jubb argues, it is plausible that an agent can be liable for a proportion of the costs of a collective harm only if i) they were in control of whether they contributed, and ii) their liability is proportionate to the relative size of their contribution. If an agent can be said to be liable for a proportion of the costs, it is 'because they could have avoided contributing at a cost at least lower than the costs for other individuals involved in an alternative distribution of responsibility that they are liable for part of the costs of the harm' (Jubb 2012 754).

Control is a graded concept. It is plausible that an agent has at least diminished control over her contribution to a collective harm if the cost of refraining from contributing to that harm would be greater than the costs imposed upon others by an alternative distribution, for example the distribution on which the costs of her contribution are allowed to fall where they lay. Given that the cost an individual contribution to climate change imposes on others is negligible (even if it is greater than zero), and given that the costs of refraining from contributing are appreciable (for example, one might miss out on the pleasure of driving a car for fun), an agent cannot obviously be said to bear outcome responsibility as a result of her contribution to climate change. This gives rise to the responsibility gap, as if every individual contributor to some significant amount of climate change-induced harm invokes the same excuse, then no one is responsible for climate change. Even if it looks too permissive to waive responsibility on the grounds that the costs of avoiding contribution are proportionately large, it is undeniable that there are a large class of cases in which the costs of refraining from contribution would be great in absolute, as well as relative terms: cases in which the agent is deeply embedded in economic structures that make it almost

outcome responsibility. Outcome responsibility, in the sense I wish to use the term, is a ground of liability (arguably the main and most important ground), though it is possible to be liable for some outcome without being outcome-responsible for it - as an insurer.

impossible to avoid contribution (of which more below). A responsibility gap is particularly manifest with respect to cases such as these.

As already intimated, we are typically interested in assignments of outcome responsibly because of their implications for questions of cost distribution. When we say so-and-so is outcome responsible for some harm, we mean there is a *prima facie* case for assigning the duty to repair that harm to her. There may be countervailing considerations that militate against assigning the whole cost, considerations of equity for instance. It may be unjust to impose a cost upon someone if doing so would bring them below some absolute level of deprivation. Nevertheless, as a general principle is often more reasonable that bearers of outcome responsibility for a certain negative impact should be regarded as bearing remedial duties in relation to that impact. Thus, we have a third framing of the collective harm problem, according to which there is a troubling gap between the quantity of harm to which remedial duties should apply, and the quantity of harm which individual agents can justifiably be required to remedy.

Stephanie Collins (2018) refers to this problem as a *collective* duty gap - collective because as climate change is caused by a group of agents, there is a kind of *prima-facie* remedial duty to repair this harm at the group level. This talk is metaphorical, though, as the group of polluters is a non-agent group, and thus is not the sort of thing to which remedially duties can coherently be assigned. The problem, viewed on this framing, is how to justify the assignment of remedial duties to sufficiently many individual agents to discharge the putative group-level duty. As we have seen, the remedial duty gap exists because each individual can deny she bears any remedial duty, on the grounds that discharging such a duty would represent a real cost to her, while her contribution to harm represents only (at most) a negligible impact upon others. Collins argues that the collective duty gap framing represents a more 'tractable' problem than the responsibility gap framing, as we can

generate individual-level remedial duties without individual-level responsibility attributions. This is because, while it is not usually possible to assign outcome responsibility to an individual on the basis of what other individuals have done (because of the control considerations just discussed), it is possible to assign remedial responsibility to an individual for what others have done (see Chapter 4).

The Structural Injustice Framing

There is a closely associated literature on group-caused wrongs, one that views them under the rubric of social-structural injustice. This tradition can be traced back, in one direction at least, to the large literature that emerged in response to John Rawls's claim that social structure was 'the subject of justice', which is to say that justice should be considered a property of institutional frameworks, not of relations between individuals. Rawls effectively identified such institutional frameworks with states, but authors that followed him argued that other types of social structure should also be regarded as falling under similar normative standards. These structures included international institutions, the system of international law more generally, and transnational economic structures like networks of international trade. When justice is viewed as a property of state institutions, justice-based critiques have a clear target: Rawls's theory of justice as fairness was effectively a handbook for people in positions of power, giving them instructions, at a high level of abstraction, as to how state institutions should be set up. As soon as we extend our conception of social structure to include, for example, structures of international commerce, the question of the target of justice-based critiques becomes more problematic. Who, apart from its victims, ought to be concerned about the existence of structural injustice? Whose problem is it?

The connection between the question of responsibility for structural injustice and the problem of collective harm should be reasonably clear. Structural injustice occurs when a system of rules or norms which govern the distribution of power and resources

systematically disadvantages one social group over another (see Haslanger 2012, Young 2011). In other words, people acting within accepted norms together contribute to producing circumstances in which certain groups are systematically disadvantaged. For structural injustice to occur, no individual need wrong any other: in fact we may suppose that each individual treats everyone else with perfect civility. Nevertheless, one is left with the impression that the victims of structural injustice have been wronged. Iris Marion Young tells the story of Sandy, a single mother who is forced out of her rental apartment by property developers, and is unable to find an apartment near enough to her place of work to it reach by public transport, compelling her to purchase a car on finance, meaning she is unable to put down the advanced rent deposit required for even a wholly unsuitable apartment, and is forced into homelessness (Young 2011 43). The people she interacts with – the developers, the letting agent, the car salesperson, the landlord - all behave generously to Sandy within the constraints placed on them by their social positions, but this is not enough to prevent her from becoming destitute. Though Sandy is the victim of injustice, it is not obvious who, if anyone, can be regarded as responsible for her condition.

Climate change can arguably be viewed as a structural injustice in the sense just outlined. The norms which make it permissible to engage in carbon-intensive activities systematically benefit people in rich countries and disadvantage those in poor countries. People in rich countries disproportionately consume fossil fuels and enjoy the attendant advantages in terms of economic development, while people in poor countries have a much higher degree of vulnerability to climate impacts as a result of having fewer surplus resources that can be put towards adaptation. It is an under-acknowledged point in the literature that uses climate change as an example in a more general philosophical debate around collective harm that the risks of climate change arise from a nexus of meteorological and pre-existing social factors. According to the IPCC, 'climate change is not a risk per se; rather climate

changes and related hazards interact with the evolving vulnerability and exposure of systems and therewith determine the changing level of risk' (Oppenheimer et al. 2014 1050).[5] This comes as no surprise the theorist of structural injustice: just as Sandy's homelessness was not caused by any one factor, nor will the condition of the 143 million people from South America, Sub-Saharan Africa and South-East Asia the World Bank estimates will be displaced by climate change before 2050 be caused by any one factor (see Rigaud et al. 2018). When one acknowledges that environmental harms are conditioned by this interaction between climate, vulnerability and exposure, and that each of these factors is itself conditioned by human behaviour in complex ways, attributions of outcome responsibility quickly start to appear inappropriate. Yet to abandon all consideration of outcome responsibility seems to let polluters off the hook for egregious injustice. Theorists including Young conclude that the recognition of considerations of structural injustice should lead us to abandon the search for outcome responsibility, but I will suggest this is a mistake (see Chapter 5).

Climate Change as Collective Harm: Limitations

It is important to note the ways in which these framings - the paradox of practical reasoning, the responsibility gap, the collective duty gap and the problem of responsibility for structural injustice – *fail* to describe climate change. The assignment of outcome responsibility might not be a problem – at least, not a philosophical problem - with respect to a great deal of climate change's impact. According to research carried out by Richard Heede, 914 $GtCO_2e$, or 67% of cumulative worldwide emissions of industrial CO_2 and methane between 1751 and 2010, can be traced back to 90 "carbon majors" – the largest companies involved in the extraction and sale of oil, coal and natural gas, as well as the

[5] I am indebted to unpublished work by Megan Blomfield for raising this point, and for this citation.

production of cement (Heede 2014). Most of these companies either still exist, were absorbed into entities that still exist, or were owned by entities that still exist (many are, or were, state-owned companies). The top five are Chevron, ExxonMobil, Saudi Aramco, BP and Gazprom (Ibid.).

At the same time, the science of detection and attribution - still in its relative infancy but improving all the time - has opened the door to the ability to link particular impacts to anthropogenic climate change. Thus, for example, (Otto et al. 2017) describe a method for estimating the impact of particular quantities of emissions, attributed to different countries or regions, on a particular weather event, a method which they demonstrate with reference to the Argentinian heatwave of summer 2013-14. They model two ensembles of possible summer temperatures for Argentina for the summer of 2013-2014, one with conditions as observed, and one counterfactual estimate for conditions without climate change. From these they estimate the change in the frequency of this event attributable to the GHG emissions of individual countries or regions. For example, they estimate that the impact of the EU's emissions made the heatwave a 1-in-12-year event rather than a 1-in-15-year event.

What both these developments mean together is that it may be possible to attribute a degree of responsibility for the harms arising from specific weather events not just to specific countries, but also to specific corporations. Indeed, not only might we be able to assign a degree of outcome responsibility to these entities, it may even be possible to attribute legally actionable liability for damages to property and loss of livelihood to these companies. Some preliminary attempts to mount this legal strategy are already under way. In the United States, the State of Rhode Island has filed suit against 21 fossil fuel companies including Chevron, BP and Royal Dutch Shell, on several legal grounds, including the impairment of public resources held in trust by the state, public nuisance and negligence for failure to warn (*Rhode Island v. Chevron Corp.*, 2019 WL 3282007 (D.R.I. July 22,

2019)). In addition to the Rhode Island suit, as of June 2019 there were 14 ongoing suits against carbon majors brought by US counties and cities (Hook 2019). Many of the complaint filings in these cases cite the very advances in attribution science just described. The prospect of holding carbon majors liable for emissions greatly reduces the gravity of the "responsibility gap" intuition with respect to climate harms. If up to 67% of the harms of climate change can be attributed to a number of entities which have legal personality and thus can bear legal liability, and with it the obligation to make whole any injured parties, then the proportion of climate harm that is in that sense morally unaccounted for is much smaller than it may first appear.

With respect to the "remedial duty gap" intuition, it is worth noting that there are a number of ways of justifying attributions of remedial duty that do not depend on the attribution of anything like outcome responsibility. Most obviously, there are considerations of ability to pay: rich states, in particular, have the ability to prevent a potentially vast quantity of climate-induced harm, at a cost which is thought to be proportionate to the amount of harm they will prevent. Thus, perhaps in an ideal world there would be no remedial duty gap, as rich states would between them distribute the burdens of responding to GHG in proportion to their ability to do so. That said, of course, this is not the world in which we in fact live. There has been a manifest and persistent failure of states to organise mitigation and adaptation efforts at a scale approaching what would be necessary to prevent the majority of the impact predicted on current emissions pathways. The question of whether there exists a justificatory route from outcome responsibility to remedial responsibility thus remains relevant, though we may still hold out hope that it will one day be rendered irrelevant by an adequate level of recognition of ability-based remedial duties.

Finally, it should be noted that considerations of individual responsibility should not be seen as the only game in town with respect to climate responsibility. It has become widely

recognised that the narrative according to which climate change is the product of individual choice, and can only be resolved when people individually decide to change their consumption habits, is one that has been systematically promulgated by organisations with vested interests in the fossil fuel economy, in order to protect those interests and to insulate themselves from reproach. The fact that, as I shall argue, there are strong reasons to think individuals have a duty to divest from certain carbon-intensive economic structures insofar as they are able, does not mean that fulfilling such duties absolves one from the responsibility to engage in political action and collective action in civil society. We ought to do both, and it is not my objective here to come down in favour of one side at the expense of the other. My aim is simply to elucidate the nature of the individual's responsibility and the content of her remedial duties with respect to climate change, such as they are.

Overview of Contents

Chapter 2 will set out existing approaches to the problem, from the perspective of a number of standard moral theories, as well as some innovative approaches intended to engage with the problem specifically. The argument generally proceeds by examining one or two typical exponents of the approaches discussed, rather than attempting to be completely exhaustive. It suggests that approaches that attempt to ground responsibility for contribution in individual wrongs cannot succeed, and identifies both strengths and weaknesses in approaches that ground individual responsibility in duties that arise from membership of coordinated groups, and approaches that ground them duties as a members of uncoordinated groups, concluding that the most promising strategy is one that can integrate these two approaches. Chapter 3 takes up a closer examination of three of the more promising recent arguments in support of the claim that individuals have direct duties to other individuals to reduce their GHG emissions, arguments based on the good of helping,

on being an essential member of a group that jointly triggers some climate harm, and on the claim that the individual directly harms others through her emissions. That none of these arguments succeed provides final evidence that no individualist approach is capable of responding to the problem.

Chapter 4 begins the "positive" portion of the argument. It defends the claim that individual obligations with respect to GHG emissions arise when the individual regards herself as involved in structures of group coordination that instantiate some, but not all, of the features of shared agency – structures which are dubbed proto-shared agency. Two accounts of potentially relevant forms of proto-shared agency in the existing recent literature are raised, and while their merits are acknowledged, they are ultimately shown to be unable to bridge the responsibility gap in the way that we need. I argue that Christopher Kutz's concept of quasi-participatory intention is, with some amendments, exactly the form of proto-shared agency we are looking for to enable us to bridge the responsibility gap. Chapter 5 takes a step back, and provides a further defence of the appeal to quasi-participatory accountability by presenting it as a preferable alternative to views according to which our existing moral concepts are judged fundamentally unsuited to guiding to individual right action and individual responsibility in the face of climate change, and in need of radical revision. It argues in favour of an understanding of the functional role of morality judgments inspired by the work of P.F. Strawson, and mobilises this view in vindication of quasi-participatory accountability in the face of radical revisionist alternatives. Chapter 6 is something of a coda: it raises the possibility of another potential source of individual moral duties in the face of climate change, duties which would be incumbent upon climate activists in particular: duties to avoid hypocrisy. It sets out a novel account of a wrongful form of hypocrisy in political discourse, and concludes that climate activists are, in all but a few atypical cases, innocent of charges of hypocrisy of this kind.

2. A Critical Taxonomy of Responses to the Problem of Individual Responsibility for Climate Change

Climate change is caused, at least in part, by individuals performing actions which, prima-facie, are non-culpable. Yet these actions, taken together, produce very serious consequences. Because these consequences are human-caused, this seems to be a case of injustice. Therefore, we are apparently faced with a case of injustice for which no one is responsible. Clearly, this is a problem, insofar as climate change is a problem, and a responsibility deficit might be thought of as a factor contributing to climate change. In the Introduction, we suggested there was also a peculiarly moral problem. But is there? If we are inclined to judge that we are faced with an injustice for which no-one is responsible, is it because this judgement is the right one? Or, does the very existence of this injustice indicate that we are lacking the proper conceptual resources to make sense of the responsibility dynamic in play?

Some theorists would deny the existence of the moral problem: our *prima facie* judgment, they say, is the correct judgement. Climate change is clearly very bad, but tackling it is not in any way a question of identifying responsible parties. Climate change is to be treated as a tragedy of the commons according to which individuals cause grave negative externalities through their *legitimate* action, and it falls to institutions to incentivise against socially harmful behaviour through measures which force the internalisation of costs. With respect to unavoidable harms, governments stand in the same relation to future victims of climate change as they do to people in need of rescue in the present day: they have a duty to provide assistance, grounded in the general duty to do good. If this is true, then there is no responsibility gap: as it stands, the assignment of responsibility for climate change is

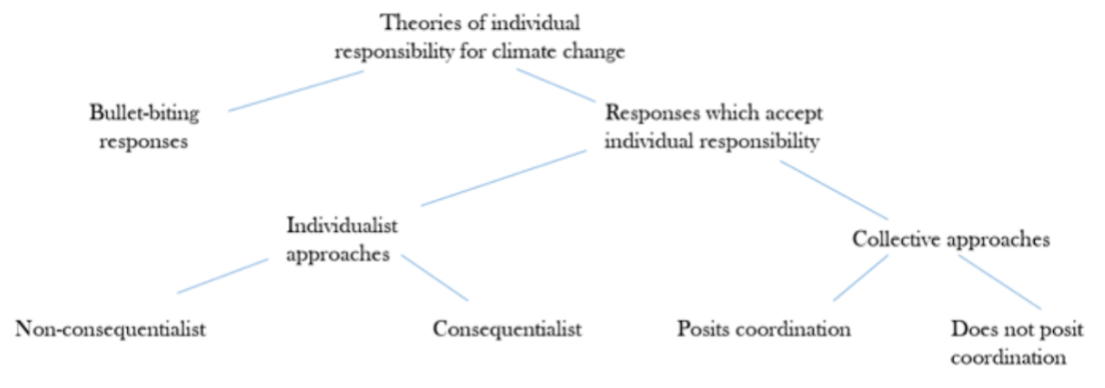
unobjectionable (some who take this view may still argue that *governments* are not doing enough to tackle the problem). I call these responses "bullet biting responses".[6]

Another approach is to argue that individuals can be held responsible for their contribution to climate change. From a consequentialist perspective, this involves demonstrating that individuals do in fact cause harm, and further that they produce more net value by not contributing than by contributing. From a non-consequentialist perspective, this involves demonstrating that the individuals' actions can rightly be condemned for reasons independent of whether they can be causally linked to harm, for example because they had malicious intent, or because they are exploitative, or because they act on maxims that fail the Kantian universalisability requirement.

A third approach is to argue that individuals can be held accountable only mediately, through their relationship with collectives or 'potential collectivities' (Cripps 2013) to which they may belong. Within this approach, some theorists assign collective responsibility by positing existing structures of group coordination, while others argue that we bear responsibility for our contribution to climate change as uncoordinated groups of various kinds.

[6] The term is somewhat inappropriate given biting this bullet is more a convenience than a cost. The thought is that the bullet bitten is the abandonment of the intuition that there is something paradoxical about the idea of avoidable human-caused harm for which no one is responsible.

This categorisation of approaches can be expressed as a tree diagram, thus:



Bullet-Biting Responses

Perhaps the central argument against a view that holds agents responsible for their emissions is some version of "we shouldn't make the best the enemy of the good". Climate change, such authors argue, is not primarily an ethical problem, but a management problem. Actors at all levels, from states to corporations to individuals, simply *will* pursue their self-interest, and in this context, the words of David Weisbach, 'there may be no room for ethics' or at the very least 'we do not need ethics' (Gardiner and Weisbach 2016 152). All relevant actors have an interest in resolving climate change, the argument goes, and we simply need to find the surest way of achieving the desired outcome from a public policy standpoint. Any considerations of justice are a separate issue which could be resolved at a later date, and if we insist on resolving the climate change problem in a way that is just, we disincentivise the relevant actors and render a resolution less likely, making the best – a just solution – the enemy of the good – a solution of some kind. This is not, however, an argument against the claim that individuals and other actors might bear responsibility for climate change, it is simply an argument against the view that we should take considerations of backward-looking responsibility into account when determining climate *policy*.

Some authors, though, have argued that considerations of responsibility simply do not come into play, even in some ideal final reckoning. A good example is Posner and Sunstein (2008), who suggest 'international paretianism' as a desirable outcome in multilateral climate negotiations (although they hold back from stating outright that they support it). This means a burden sharing agreement on which no country is worse off when the costs and benefits of the deal are netted out. It is estimated that the cost of climate change to the United States will be much less, as a percentage of GDP, than the cost of climate change to the most vulnerable states, thus it is possible that an international burden sharing agreement that is optimal from the perspective of the world as a whole will constitute a net loss for the United States. International paretianism would in such a circumstance entail side-payments from the most vulnerable countries (which also happen to be some of the poorest) to the United States, to compensate the USA for its mitigation efforts and provide an incentive to sign up to the deal.

A key strand of their argument is that considerations of corrective justice – the necessity of compensation for loss and damage - are not as relevant as is usually assumed when it comes to climate equity. Carbon-intensive economic activity, they argue, produces economic growth, which can then be drawn upon to help people in the future. So, it is necessary to consider whether the benefits produced for future people by current economic activity are greater than the costs. While these authors do not explicitly state that the costs do not outweigh the benefits, other authors have made this claim. Notably (or infamously), Bjorn Lomborg (2001) - cited suggestively by Posner and Sunstein - argued that climate change mitigation was a very inefficient means of helping the global poor and that if we adopted anything but a carbon tax much lower than one that would be necessary to keep global temperature increase below 1.5C or 2C, this would constitute a net social cost. There are a number of ways to make sense of this position. One would be to say that although the

current generation will harm people in the future if we allow global average temperature to rise above 1.5C or 2C, we compensate people for that harm through the productive capital we pass on to them. Another would be to say that although we harm future people by allowing temperatures to rise, we also harm them by failing to invest capital efficiently to maximise future prosperity, and we must choose the lesser of two evils.

Although Posner and Sunstein do not deny outright that particular actors could be considered culpable for their emissions, they doubt that this is possible to any significant degree. '[I]t is easily imaginable that the costs of emissions abatement [for a particular actor] would be significant', they write, 'it is also easily imaginable that the benefits of emissions abatement [with respect to the emissions of a particular actor]...would be close to zero' (Posner and Sunstein 2008 1599). Given that the benefits of cutting emissions are 'trivial' and the costs are significant, it is not appropriate to regard the emitter as negligent, and therefore not appropriate to regard them as culpable. On the basis of calculations by William Nordhaus, they estimate the social cost of driving a car, by way of example, to be $0.10 per gallon of petrol consumed. Negligence, they claim, is the failure to take cost-justified precautions against risk. So, they argue, if by choosing to drive rather than walking, a driver obtains a benefit worth more than $0.10 per gallon of petrol consumed (a very low bar by anyone's reckoning) they are non-negligent, and therefore non-culpable.

To summarise, they claim at least one of:

i) *Compensation*: we compensate future people for harms through capital investment

ii) *More harm than good:* aggressive mitigation would do more harm than modest mitigation

and further claim

iii) *Proportionality:* individuals cannot be held responsible for harm when the cost of avoiding it would be greater than the cost imposed on the victim(s), and this condition is satisfied for most or all relevant actors with respect to greenhouse gas emissions.

Each of these claims can be resisted; I give a sketch of how this can be done, below.

i) *Compensation*

In Dickens's *Hard Times*, Sissy Jupe recounts that the schoolmaster Mr. M'Choakumchild told her the gross domestic product of her nation, and demanded she answer as to whether she were living in prosperity. She reasoned, 'I thought I couldn't know whether it was a prosperous nation or not...unless I knew who had got the money, and whether any of it was mine' (Dickens 1854, ch. IX), and was roundly rebuked for her stupidity. If, unlike her teacher, we are inclined to think she was onto something, then in much the same spirit, we might judge that it is rather implausible to claim that the victims of climate change will be compensated by the higher level of economic growth promoted by a cheap carbon price, until we know who is benefiting from the growth, and who is being harmed by climate change. Given that Posner and Sunstein's argument is actually premised on the fact that the USA stands to lose a fraction, in GDP terms, of what a country like India stands to lose, it seems strange not to question whether a future Indian will be compensated by the economic growth generated by an American company. As a much-publicised report by Oxfam noted, the poorest half of the world's population has received just 1% of the total increase in global wealth since 2000, and since 2010 their wealth has actually fallen by 44% (Ayele, Fuentes-Nieva and Hardoon 2016 2). There is no reason to suppose that future growth will be any different – we have no reason to believe that the gains of economic

growth will be distributed evenly across the globe, or even that all regions will be better off than they would otherwise have been if certain regions are allowed to grow at a faster rate.

As well as compensation failing because it does not reach the victims, it may also fail to repay like with like. Perhaps the most worrying metric on which climate change will cause harm is the effects it will have on health outcomes (see Costello et al. 2009). As Wilkinson and Pickett (2009) have forcefully argued, inequality is itself a determinate of health, irrespective of peoples' absolute level of wealth. Thus if the current trend whereby economic growth has favoured greater wealth disparity continues, it is possible victims of climate change may actually be rendered *less* capable of adapting as a result of carbon-intensive economic growth than they would be if the same emissions took place without the growth, if the health benefits of growth are outweighed by the health cost of inequality. Should we consider this a likely prospect? Though it is certainly true that there is a correlation between carbon-intensive growth and positive health outcomes, it is not a necessary relationship. Cuba produces health outcomes on a par with those of the United States, with less than a fifth of its per capita ecological footprint (Wilkinson and Pickett 2009 221). Thus at the very least we can deny it is the case that we *need* carbon-intensive growth to guarantee future health outcomes, and it is possible such growth even takes us further away from the goal of ensuring victims are 'compensated'. This brings us to the second point.

ii) More harm than good

In determining whether aggressive mitigation would do more harm than modest mitigation, consider a possible scenario. Say individual emitters choose to reduce emissions today, and as a result, in thirty years' time, GDP is 3% lower than it would have been if they had chosen

a less aggressive mitigation strategy.[7] In consequence (sticking with the same metric of harm for convenience's sake), health outcomes are poorer than they would otherwise have been. For example, life expectancy is lower. Imagine now a second scenario, on which those emitters do not adopt an aggressive mitigation strategy, and as a result of the climate change caused, an entirely separate group of equal size has a life expectancy that is lower than it would have been if those emitters had adopted the aggressive mitigation strategy, and by exactly the same margin as in the first case. Do these two costs in wellbeing cancel out, such that a single individual made worse off on one side rather than the other would provide decisive reason to choose that policy?

There are a number of ways we could distinguish the two. Perhaps one counts as genuine harm, while the other does not. Perhaps, while both count as harm, only one counts as wrongful harm. Or we perhaps might appeal to a distinction between doing and allowing harm. *Ex hypothesi*, there is a direct causal relationship between the lower life expectancy in the climate change case, and the emissions (in practice this would be very difficult to establish but let us suspend our disbelief for the moment). Say for example a warmer climate, caused by emissions, led to more mosquitos and a higher incidence of malaria. Plausibly, there may not be the same kind of causal relationship on the GDP loss side: perhaps it is a result of less research funding, meaning that a new cure is not discovered for a particular prevalent disease. In one case, the actions of emitters caused the incidence of disease to rise, in the other, the incidence of disease remained the same, but a new means of rescue never emerged. We might think this is a relevant distinction.

One way that we can make sense of the natural judgement that this distinction makes a

[7] In the context of Posner and Sunstein's argument, this would be done through a carbon tax, rather than spontaneous behavioral change without some market mechanism. In setting out this scenario, however, I want to avoid making the change a result of state policy to avoid the non-identity problem, and because the primary focus of this thesis is individual responsibility.

morally significant difference is through noting the degree to which each of the impacts is mediated through human agency. That health outcomes are poorer "as a result" of lower economic growth depends on a number of decisions by human agents: the rates and methods of taxation, the rate of public sector investment, the particular investment decisions of various governments and agencies and the behaviour of other market actors, to name but a few. On the other hand, the processes through which carbon-emitting actions cause climate change are mechanistic. Beyond the assumption that people in other countries will continue to emit carbon, the causal link between emissions and poor health outcomes does not depend on other people exercising their own agency in some particular ways, downstream in the causal chain. The fact that these downstream actors have full control over whether the outcome occurs casts doubt on whether economic activity now can be considered a cause of better health outcomes at some point in the future.

A further point: if failing to maximise economic growth harms future people, and one has a duty not to produce such harm, then one would have a duty to maximise economic growth to the best of one's ability. The idea individuals might have a duty to maximise economic growth looks highly implausible. Such a duty would restrict the individual's options to an unacceptable degree, rendering her essentially bound to do whatever, with her particular skills, would have the greatest impact on GDP. It might be countered that a middle ground could be found, such that one was bound to choose which ever of an acceptably large range of maximally productive careers appealed to one. Even this, however, looks worryingly illiberal and restrictive, a kind of supercharged Calvinism which is out of step with contemporary values, and certainly with the classical liberal leanings of Posner and Sunstein. A duty to minimise contributions to climate change, however, is apparently much more widely accepted. It may be that this is just an appeal to the prevailing ideology, but it is often a good guide in moral philosophy to suppose that a difference in

practice represents a meaningful moral distinction.

iii) Proportionality

On the question of what costs it is reasonable to impose in the pursuit of harm-avoidance, things get tricky. It really does look difficult to say that an individual might be culpably negligent simply for - say - driving a car from A to B as a result of the related emissions. In the following sections of this chapter, I will consider a number of responses to this problem, some of which do accept backward-looking culpability/liability, and others which draw a distinction between culpability and responsibility more broadly, arguing that although individual liability is indeed ruled out, individuals can still be responsible for their *involvement* in climate change, and this can generate certain duties.

It is, however, possible at this point to make some remarks about Posner and Sunstein's argument against culpability in particular. While a social cost for carbon may be a useful tool politically, it arguably should not be seen as buying yourself a free pass (an 'environmental indulgence', to borrow Robert Goodin's (1994) religious analogy). This is partly because carbon price is not a simple representation of the external cost associated with those emissions; some element of cost-benefit analysis already goes into setting carbon price. As Nordhaus writes, 'the carbon tax ... balances the marginal social costs and marginal social benefits of additional emissions' (Nordhaus 2008 149). Thus it is disingenuous to assert the threshold level of benefit at which driving becomes a net social benefit as though it were an objective matter, as some evaluative assessment of what would constitute too great a loss of economic benefits is represented in the carbon price itself.

Furthermore, some impacts of climate change cannot easily be priced. Imagine, for example, an indigenous community whose island home is to be submerged by sea level rises. If price is calculated as the point at which an individual would be indifferent between

a sum of money offered and the good in question, it is plausible that for this community, no indifference point with respect to their homeland would ever be reached, and that it is in that sense priceless. An alternative method of pricing a good is determining the amount of money an individual would be prepared to pay to keep the good. The problem with this method is it is hard to view is as a mere heuristic, as we are inclined to doubt whether the imaginary questioner has the right to make such a demand.

This has only been a sketch of the various ways in which bullet-biting responses may be found wanting. More certainly needs to be said in order to give a full dismissal of them, but it is not my intention to do so here. These responses, while politically very influential, are heterodox among philosophers in the field of climate ethics. It is widely accepted in the philosophical literature at least that questions of responsibility are highly relevant when considering why and how carbon emissions should be reduced and regulated. What this section has suggested, however, is that the claim that there is no duty to mitigate emissions arising out of norms which militate against harming others is inadequately supported by bullet-biting theorists. Both the suggestion that economic activity now constitutes some kind of pre-emptive compensation against future harm and the suggestion that mitigation is itself a form of harm are deeply suspect.

Consequentialist Individualist Approaches

Theorists in this category hold that individuals are responsible for their contribution to global climate change, because those contributions cause significant harm. When moral theorists believe that what matters is consequences, they typically accord little importance to the distinction between the badness of doing versus allowing harm. Thus, perhaps a typical consequentialist understanding of our responsibilities with respect to climate change is to regard them as the same as our responsibilities with respect to natural disasters, or to avoidable suffering of any kind. Affluent individuals on this view have an obligation to alleviate the maximum amount of suffering they can, without sacrificing anything of comparable importance to themselves (Singer 2010 231). The view would therefore be that because climate change will cause mass suffering, the individual has some duty to help the victims of that suffering, but reducing her emissions is not the most effective way of doing it. The best way may be devoting her time to lobbying government, or donating to climate adaptation funds. If she can do these things more effectively while not reducing her individual emissions, then on this view, she has no obligation to do so.

Some theorists, however, have argued that the (broadly consequentialist) value of avoiding harm does give us good reasons to avoid contributions to greenhouse gas emissions. John Broome, perhaps the leading exponent of this kind of approach to the problem, argues that individuals cause harm through their emissions, that this harm should be regarded as an injustice, and that individuals are culpable for causing injustice.[8] He gave a statement of

[8] This language of "injustice" is not incompatible with consequentialism on a broad definition, which is how Broome himself, for example, understands the term. Although Broome's argument has features associated with non-consequentialism, such as an apparent appeal to the distinction between acts and omissions, it may still be regarded as consequentialist in that it evaluates the rightness of acts only in terms of the goodness of states of affairs (see Broome 1991 1-20). A very large number of considerations are taken to count towards the goodness of states of affairs, however, including considerations of justice. The form of consequentialism Broome adheres to allows for agent-relative reasons where others do not. This is why, for example, the doing-allowing distinction is salient on this view, where for consequentialists who only acknowledge agent-neutral reasons, it may not be.

this view in *Climate Matters* (2012) and has defended it several times since, including in (Broome 2016; 2019). As stated in *Climate Matters*, he gives seven reasons. Broome claims neither that any of these conditions is necessary, nor that any subset of them is sufficient for regarding individual responsibility to climate change as an injustice. It is therefore somewhat puzzling what role they are supposed to play. One must suppose that they are intended cumulatively to provide support for the claim. They are: i) the harm is the result of an act (rather than an omission), ii) it is serious, iii) it is non-accidental, iv) it is not compensated, v) the emitter benefits from her harmful activity, vi) the harm is not reciprocated, and vii) the harm could easily be avoided. We are under a duty to refrain from committing injustice, and if we do, we are required to make restitution. Thus in rich countries at least, individuals have a duty to reduce their emissions to zero. The claim that they incur liabilities if they fail to do so might also be thought to be implied.

One might resist the idea that any of Broome's conditions in fact hold, but i), ii) and vii) are probably the most controversial. Speaking more broadly, these conditions correspond to three problems that seem to apply to individualist consequentialist approaches in general, which we may call the causation problem, the aggregation problem, and the legitimate excuse problem.

With respect to i), we must remark that it is difficult to establish the existence of a causal pathway between particular emission-causing acts, or even an individual's cumulative emissions over a lifetime, and particular harms. According to Broome, as a general rule 'the harm done by greenhouse gas emissions is proportional to the quantity of emissions' (Broome 2016 160). So calculating how much harm each individual does should be a simple matter of calculating their emissions as a percentage of total emissions, and taking the same percentage of total harm. Broome admits, however, that 'at the much smaller scale of, say, thousands of tonnes, the harm is lumpy' (Ibid.). On the figure Broome cites

(attributed to David Frame), the average person in a rich country will emit roughly 800 tonnes of CO_2 in a lifetime. This, clearly, then is at the 'lumpy' end of the scale, by Broome's own lights. By lumpy, Broome means that a graph of quantity of emissions against quantity of harm would not describe a straight line or a smooth curve, but would be stepped, such that some increases in emissions would correspond to increases in harm, while others would not. Given that we cannot know where these plateaus and jumps lie, because the climate system so complicated, we are entitled, he claims, to smooth them out. We are then talking about expected harm.

The shift from talking about harm to expect harm casts some doubt on Broome's claim that the harm is a result of an individual's act. To say that one's action has some expectation of harm attached to it is not the same as saying harm results from one's action. Even causing harm is not a sufficient condition for injustice, so we may think that causing expected harm has even less of a claim to be a significant factor in demonstrating the presence of injustice. Exposing others to a certain level of expected harm is generally considered unobjectionable. The ethical question then shifts from whether one is bound by a norm against harming others, to whether one is taking enough precaution against risk. Broom would have to show that the level of risk to which the emitter exposed others was unacceptable: this is precisely what the bullet-biters deny. If, as Posner and Sunstein claim, I am culpable when I fail to take precautions against risk that are justified by the expected cost, the consequentialist argument would need to establish this proportionality between the size of the expected harm and the cost of avoiding it. This is presumably why Broome acknowledges the central importance of condition vi), that harm can be "easily" avoided.

With respect to ii), a commonly discussed problem is whether it is possible to say that through my emissions, I do a very small amount of harm to a very large number of people, and if so, whether it is possible to infer from this premise that I do serious harm. But there

is a question prior to these: what do we mean by serious? If we mean, harm that we have all-things-considered reason to avoid, what are the criteria here? If 'seriousness' means something like 'aggregate weight of negative expected value', is there always some level at which 'serious harm' becomes morally significant? Is that level reached with respect to the emissions of the average individual?

Perhaps "morally significant harm" should designate some substantive criterion like the violation of a right or the frustration of a fundamental interest. If so, it looks like it is going to be very difficult to establish "significant" harm is perpetrated in the case of an individual's contribution to climate change. Even if we accept that several imperceptible harms add up to serious harm, it's not the case that several imperceptible harms that do not constitute rights violations add up to a rights violation. So it looks like when Broome claims we do 'serious' harm, this does seem to be constituted by nothing more than a sufficiently high level of expected aggregate harm on some undifferentiated metric. We may doubt that this condition is necessary or sufficient for the presence of an injustice.

Then we have the aggregation problem itself: the relationship between individual contributions to climate change and harm is arguably a case of a sorites paradox, where it looks like individual contributions do not make a difference to harm, although those contributions add up to significant harm. There have been two main types of response to this problem. Following Parfit (1984 77), consequentialists might claim that each individual does imperceptible harm, and we should regard such actions as worthy of moral condemnation for the very reason that a lot of imperceptible harms add up to serious harm. Even accepting that the individual *does* do imperceptible harm in this case, though, this approach arguably fails to provide an explanation of *why* merely being the cause of imperceptible harm should be morally culpable, especially when the actions in question produce clearly perceptible benefits.

Supposing we do not accept that the individual does imperceptible harm, another option is to argue that because there is a significant probability that an individual will trigger a threshold for some sufficiently serious harm, the individual does significant expected harm (see Kagan 2011). The problem here is the difficulty in showing that the expected value calculation does indeed come out negative: we would need to see evidence that the threshold harm is sufficiently serious and the probability of triggering it sufficiently great that it outweighs the expected benefit of fossil fuel use, and it does not seem that this has been satisfactory shown as of yet (see Nefsky 2011).

It might be argued that it is not difficult to show that the individual does enough harm or expected harm to be significant, so long as it is the case that the harm could be easily avoided. This brings us to consideration of Broome's point vii). This condition is cashed out for Broome by the assertion that we can easily avoid causing harm by paying for carbon offset. This means that we can purchase certificates under various certified schemes, which reduce the amount of carbon going into the atmosphere in a number of ways, including by investing in renewables, investing in efficient infrastructure for development, planting trees and reducing methane from landfill. Purchasing offset clearly has certain a cost attached: a brief survey gives prices ranging from about $8/tonne to about $18/tonne.[9] If one emits 800 tonnes CO_2e in one's lifetime, this would work out at about $200 a year over fifty years of working-age life. It is therefore easy to avoid emissions? In countries such as the UK, it is probably relatively common for individuals to spend the equivalent of $200 a year on something like coffee, so it is "easy" to avoid emissions in the sense that most people could afford it if it was necessary. On the other hand, $200 a year is an appreciable sum. To respond to the bullet-biters' challenge, Broome would need to show that this sum were proportionate to the level of risk the emitter would otherwise impose. Broome estimates,

[9] See for e.g. https://www.carbonfootprint.com/offset.aspx?o=10, [accessed November 12, 2017]

on the basis of WHO figures and the David Frame figure, that over the course of a lifetime, the emissions of a person in a rich country will 'wipe out' over six months of human life (Broome 2012 74). This odd locution fails to obscure the fact that 'wiping out human life' must in this context mean something quite different from shortening anyone's life to that extent, or killing anyone. The figure of sixth months is generated by taking the ratio between individual emissions and total emissions, and applying this to the total amount of harm, to derive "individual harm". But being one of n people who together cause a harm in some way, does not automatically make me responsible for $1/n$ of that harm – this would seem to an instance of what Parfit called his first mistake in moral mathematics (Parfit 198468).[10] Thus it does not appear that Broome has the resources to estimate whether a $200 cost is proportionate to whatever risks the individual in fact imposes, and thus whether they can be considered culpable if they do not spend this money.

More importantly, however, Broome's argument assumes carbon offset schemes are viable, and that they would not involve us in new patterns of collective harm which may equally constitute injustice, or that they might not constitute forms of injustice in themselves. If such schemes are not viable, then it is untrue that one can easily avoid causing harm: for an individual to divest from fossil fuels would involve significant cost. For rich individuals, it may be possible, but would most likely involve sacrificing fundamental interests (it would require one to return to a kind of pre-industrial existence). For poor individuals in industrialised societies, it would probably be impossible without losing the bare necessities of subsistence. Such costs would be far from trivial. The claim that individuals are

[10] Parfit characterises as mistaken the "share-of-total view", according to which if n people together produce X units of benefit, each person can consider herself responsible for X/n units of benefit. Parfit points out this is absurd in circumstances in which n - 1 people would also have been enough to produce X units of benefit: superfluous people should not be credited with producing benefits when they could non-superfluously help to produce benefits elsewhere. 'Similar claims apply to harm', Parfit writes (Parfit 1984 69).

responsible for their emissions because they cause significant, avoidable harm is thus by no means a straightforward matter.

Non-Consequentialist Individualist Approaches I: Kantianism

Formula of Universal Law

Kant's formula of universal law is often viewed as articulating the same principle as the admonition "what if everyone did the same as you?". Given this reading, one might think that a Kantian approach would be very well suited to dealing with the problem of collective impact, as this is precisely a problem that arises out of many people performing acts of the same type. When we get down to the specifics of Kantian theory, however, we see that it is not nearly so simple. Commentators including Allen Wood have argued that Kant's Formula of Universal Law is unworkable as a test for the wrongness of acting on maxims, as on any reasonable interpretation it admits of many counterexamples (Wood 2011). The argument in this section therefore has two strands. First, I argue that, on interpretations we can regard as reasonably plausible and reasonably authentic, Kantian arguments do not condemn individual contributions to the collective harm of climate change. In this, they are at least no more defective than individualist consequentialism: we cannot say without further argument that they give the *wrong* result, but we can say that they are unhelpful for our purposes. Second, I will point to interpretations of Kant on which, in cases that are closely related to the problem of collective harm, results are generated which determinately conflict with widely accepted moral principles. This adds to our suspicion that they will not be successful in providing a solution to the problem.

In what follows, I do not defend a particular interpretation of Kant's ethical system, but rather present a generalised picture, which draws mainly on Christine Korsgaard, and also consider Derek Parfit's "improved" version of the theory (Korsgaard 1996; Parfit 2011a). I have elected not to treat Kant's text directly, for two reasons. First, a multiplicity of interpretations are available and it falls beyond the scope of this project to defend a particular interpretation from first principles. Second, because I wish the theoretical

content to be as clearly defined as possible, ensuring the discussion is focused on the application of the theory to the problem of collective harm, and not on interpretive debates. While it may be that certain other interpretations have the resources to defend themselves against the criticisms I adduce against these accounts, which I have not considered, I believe the critique is sufficiently generic that, even if it not immediately effective against some interpretations, it can be adapted to make it so.

There are well-known difficulties with finding a coherent reading of Kant's formula of Universal Law. It is not the case that any maxim that, if followed by everyone, would produce bad effects, ought to be forbidden. There are many such maxims that are rightly considered morally innocuous. If all of a bank's customers demanded their balances in cash at the same time, for example, this could have disastrous effects, but that does not mean that it is wrong for any individual to empty her account in normal circumstances. It would therefore be uncharitable to suppose that the universalisability test is supposed to function as banning any act that would have bad consequences if everyone did that act at the same time, and alternative readings bring this out.

On Christine Korsgaard's interpretation, Kant's formula of universal law prohibits us from acting upon a maxim that, were everyone to act upon it, the purpose specified in the maxim would be frustrated (Korsgaard 1996). So one may not act on the maxim "I will assassinate my rivals to gain advantage", not because a world in which everyone acted on that maxim would be worse than a world in which no one acted upon it (although it would be), but because in such a world, the agent's purpose of gaining advantage for herself would not be fulfilled. To enjoy any advantage requires that one can retain that advantage securely, and this would be prevented because one would be vulnerable to assassination oneself. The stipulation that a maxim must specify a purpose for action overcomes the bank run problem, because not everyone will have the same ends in view at all times. A maxim is to

be thought of as a standing policy, which takes effect only when a particular end comes into view. The relevant maxim here would be something like "I will empty my account in order to switch to a different bank". This maxim is universalisable, as only a limited number of people will want to switch banks at the same time.

Now, when one contributes to greenhouse gas emissions, presumably, one is acting on maxims like "I will drive cars to get to my destination quickly and comfortably, whenever I have no reason to do otherwise". A world in which this maxim is a universal law would presumably look very like the world in which we actually live. In this world, I seem to be quite capable of achieving my purpose: reaching my destination in comfort. The same seems to be true for any maxim which might describe the policy I am following for any carbon-intensive act type. If this is right, Kant's first formulation of the categorical imperative does little to help us solve the problem of collective harm in the case of climate change – it under-generates.

The Formula of Universal Law seems to work best when applied to cases where the wrongness of some action involves the idea that it exploits some generally useful convention for one's own purposes. This is why Kant repeatedly adduces the example of making a lying promise in order to benefit oneself: it seems plausible that what is wrong with such an act is that it relies for its success on the convention of honesty, and that one defects from the convention of honesty for one's own benefit – one implicitly endorses that convention and the benefit one extracts is made possible only by the compliance of others with it. In order for cases of contribution to collective harm to have this structure, therefore, there must be some convention in place, upon which individuals can be considered to be freeriding, some "system" they are "gaming". When no such convention exists, it looks difficult to say that anyone has done anything wrong by contributing. Some authors have argued that in collective impact cases, we should act as though such conventions were in

fact in place. Christian Baatz, for example, has argued that we have a *prima facie* duty to reduce our emissions to the level that would constitute a 'fair share' of global absorptive capacity for GHGs (Baatz 2014). Acknowledging structural dependence on carbon-intensive economic systems, he argues that individuals should be regarded as having a Kantian imperfect duty to reduce their emissions insofar as they are able, within the confines of the economic structures in which they find themselves.

There is a difference, however, between there being a case for people to be assigned duties, and their actually having them. In law, a distinction is drawn between *malum prohibitum* and *malum in se*. Where acts that fall under *malum in se* are "natural wrongs", acts that fall under *malum prohibitum* are only wrong once a regulatory convention has been established. John Gardner applies this distinction to discrimination law (Gardner 2018). In a society pervaded by racism, racial discrimination can be regarded as a collective action problem. White businesses owners, for example, may not have explicitly racist motivations, but may be unwilling to serve customers of colour for fear of losing their white clientele and thereby their livelihoods. Gardner argues discrimination may not initially constitute wrongdoing on the part of the business owners, because they act for good reasons. But the hugely detrimental impact of discrimination on communities of colour does justify a duty on the part of lawmakers to legislate against it. Only when the law is in place does it become wrong for business owners to refuse to comply with it, because it is only when the law is in place that the weight of reasons in favour of non-discrimination becomes decisive. Similarly in the case of climate change, though there is a good case for regulation, and though appropriately placed regulatory bodies may have a duty to intervene to institute rules to curb carbon emissions, it is contrary to the norms implicit in jurisprudential practice to suppose that an individual is required to behave as though some appropriate regulation were in place.

Kantian reasoning, then, does not provide a case for an individual duty not to contribute to collective harm grounded in the wrong of misusing a valuable convention. What would be worse, Kantian reasoning may seem to give us what is intuitively the *wrong* answer in collective harm cases – specifically, cases in which a certain act type is unobjectionable if few people do it, but which we want to condemn if many people do it. Parfit has called this the threshold problem. I will argue that Kantians can defend themselves against this accusation, although this does not take them closer to providing an adequate response to the problem of collective harm for the reasons already adduced. In response to the threshold problem, Parfit introduces his own "improved" version of the formula of universal law, which he believes gets much closer to giving the right answer in cases of group-caused harm. I will argue that he is mistaken in this view: his "improvement" does no better.

Take the classic example of a "tragedy of the commons": a small community is sustainably fishing in a lake, where the fish in the lake constitute a common pool resource, characterised by a depletable stock and a flow of renewable consumable goods. Now, the number of people in the community suddenly expands. If everyone were to act on the maxim, "I will fish in the lake to get enough food to eat", then fish stocks would be depleted, and the end of getting enough food to eat would not be achievable by this means. Nevertheless, it looks like the original sustainable fishing group was doing nothing wrong. Baylor-Johnson (2003) suggested that, for this reason, Kantian theory gives the wrong result in a tragedy of the commons (although he gives no detail as to his interpretation of Kantian theory and precisely how it conflicts with common-sense morality in this case; he modified his view in Baylor-Johnson (2011)).

A reply is available to the Kantian: that a maxim must specify circumstances in which the agent acts, and this specification will state whether the agent is in circumstances in which

overfishing is possible, which will prevent over-generation. Thomas Pogge (Pogge 1997) advocates this approach. Parfit suggests that the proposed solution fails because it would 'make our moral reasoning take a rather strange form' (Parfit 2011a 210). Take the maxim "I will become a dentist".[11] The Formula of Universal Law seems wrongly to condemn this maxim as, were it universalised, the job market would become flooded with dentists, causing a threshold effect whereby dentists can no longer find employment. So, the suggestion under consideration goes, we add the caveat "...in circumstances in which there are not already too many dentists". Parfit objects that the universalisation of this conditional maxim would still lead to a world in which everyone aspired to become dentists and were disappointed.

Even if we accept Parfit's interpretation of the Universal Law formula, however (of which more below), it looks like this objection misses the mark. All the Kantian formula is supposed to do is test whether acting on certain maxims is permissible. It says nothing about people's desires. I may hold the maxim "I will become a dentist, to earn a secure living, as long as there aren't too many dentists on the job market", as a background policy, which simply doesn't enter into my day-to-day consideration when the condition does not hold. I have decided dentistry is one good means of earning a secure living, under the right conditions, but there may be many others. Similarly in the fishing case, a world in which the maxim 'I will fish to get enough to eat, as long as too many others are not doing the same' were universalised does not necessarily resemble a world of frustrated would-be

[11] Here I adopt Parfit's use of the term maxim, although I do not believe that this can straightforwardly be regarded as a maxim as it does not apparently specify a purpose. Although we could perhaps read the maxim as 'I will become a dentist to achieve the value of being a dentist', this seems psychologically unrealistic, it is implausible that being a dentist is an intrinsic feature of anyone's conception of the good, as opposed to, say, helping people, or having a secure living. That said, according to Parfit, the whole idea that we can generally be regarded to be acting on maxims at all is psychologically unrealistic.

anglers, as if the condition is not met, everyone will just pursue their end of getting enough food to eat in other ways.

Even if this rescues Kantians from the charge that they get the wrong result about maxims relating to the use of common pool resources, however, for the reasons already adduced, it still looks like Kantians have not got us any closer to an answer to the problem of collective harm in the case of climate change. Parfit argues for an idiosyncratic, "improved" interpretation of the first formulation of the categorical imperative that, he claims, gets us much closer to providing a solution to collective harm cases. He argues that it is most coherent to interpret Kant's Formula of Universal Law as requiring us to ask whether we could rationally will that we lived in a world where everyone acted in some way, rather than a world in which no one did (he abandons the device of the maxim) (Parfit 2011a 301). Here, Parfit means "rationally will" to be synonymous with "rationally choose": we are supposed to make an evaluative calculation between two hypothetical states of affairs (Parfit argues deontic reasons must be ruled out from this calculation, on the grounds that this would a kind of bootstrapping – it cannot be that it is wrong for me to perform some act because, were the maxim on which I acted universalised, the world would contain a lot of wrongdoing. This would leave us with reasons about goodness). At a first pass, it seems that this interpretation is ideally suited to addressing the problem of collective harm in the case of climate change: Parfit's Kant asks us whether we would rationally choose a world in which everyone performed acts which contributed to climate change, over a world in which no one did. Given that a world in which climate change occurs is worse than a world in which it does not, the principle judges that acts which contribute to climate change are morally prohibited. To be clear, it might not be worse in terms of the narrow self-interest of the agent, as she may not live to see the worst effects of climate change, and would reap

the benefits of carbon-intensive activity.[12] On the reasonable assumption, however, that each of us also attaches value to the interests of our children, our friends, etc., it is plausible that none of us could "rationally will", in Parfit's sense, that such a world be instantiated.

This reading, however, does not overcome the threshold problem. In cases in which all would benefit if everyone contributed to some public good, and everyone would lose out if no one contributed to that good, Parfit's Kant requires us to make our fair share of contribution to that good. Importantly, it is required even in circumstances of global non-compliance. On this interpretation then, a Parfitian-Kantian is required to make her contribution to a public good, even if to do so would be pointless, insofar as the public good will not be attained even if she contributes. Take some classic public good like military defence: as long as I prefer a world in which everyone contributes to the good, to a world in which no-one contributes, I am morally required to contribute to defence spending. I am required to do so even if my contribution is not enough to pay the salary of a single soldier, and no one else contributes. An advocate of such an approach might just stick to their guns by saying that this is one of many instances when it is just hard to do what is right. But if it seems unreasonable, we may well wish to reject this version of Kant's view. [13]

Applying this directly to the climate change case, the point would be that my act of failing to lower my carbon emissions would be judged to be wrong on Parfit's Kantian principle

[12] Steven Gardiner (2011) makes this point at length, pointing out that climate change is not a standard prisoners' dilemma in that cooperation may not represent the optimal outcome for all currently extant players, whether these are thought of as individual private citizens, or as political leaders, with institutionalised short-term priorities.

[13] Note also that whether failure to contribute is wrong arguably depends on the scope of "everyone". I might prefer a world in which no one *in the world* contributed to defence spending to a world in which everyone did, as perhaps a world without military funding would be one in which fewer wars were fought. Furthermore, if we took the universalised version of my maxim to be "it is a universal law that people contribute to the defence *of the UK*", I might rationally find this world less preferable, insofar as I value equality of power between countries, or what have you. When considering common-pool resources, the scope of "everyone" seems naturally limited to those people who benefit from the common pool resource. But it is not clear that specification can be found in Kant.

even when no one else will lower theirs, because I cannot rationally choose a world in which everyone refuses to lower their carbon emissions over a world in which no one does. This, Parfit argues, cannot be right, because it cannot be that I am required to make pointless sacrifices. Could we overcome this problem by adding a specification of the circumstances to the description of the act, in the manner Pogge suggests? It seems not. If everyone acted on the maxim "I will refuse to lower my carbon emissions except if a sufficient number of others are lowering theirs" then we would be left in a world in which, assuming the threshold number had not already been reached, we would be faced with insurmountable inertia.

This is in fact a generalizable point: whether we take Korsgaard's reading, Parfit's revamped interpretation, or other interpretations, it looks like even if we build the idea of circumstances into our maxim, the method gets the wrong result. It gets the wrong result, because it 'reduces other agents to background resources' and thereby 'denies the agency of all by focusing on the agency of each', to borrow the language of Christopher Kutz (2002 479). If no-one is cooperating, then emissions are permissible. If some threshold number of people are cooperating, then emissions are not permissible. We have, as it were, an impassable wall between these two potential worlds, and the Kantian formula does nothing to tell us which of the two ought to be brought about. In order to solve the problem of collective harm, it seems what is required is an ethical approach that can operate in the grey area between the permissible and the impermissible, a task for which the Formula of Universal Law seems far from ideally suited.

Formula of Humanity

Kant's formula of humanity arguably does no better: why should my act of driving a car from A to B count as failing to treat anyone as an end in themselves? In order to claim that I violate someone's autonomy, or their rights, or fail to respect their humanity in any way,

it seems to be the case that I must have a definite impact on them. I do not mean to suggest this as a generalizable description of the application of the Formula of Humanity: we might believe there are cases in which I violate the formula, where it seems reasonable to say I have no impact on the person I fail to treat as an objective end. An example might be deceiving someone from paternalism, in a way that they would choose to be deceived, were this possible, and where the concealed information is irrelevant to any of their ends.[14] For present purposes, I have no reason to deny that the Formula of Humanity might be correct to condemn such an act. My claim is rather that in the core examples of collective impact cases under consideration, it appears the only suitable candidates for reasons why I might be considered to fail to treat some other as an objective end are reasons which involve me *doing something to* them. To fail to treat someone as a source of objective ends is to treat them in such a way that prejudices their ability to set ends for themselves in some way. But it is difficult to see what end a person could have that was prejudiced by my driving of a car from A to B, unless he had some claim against me acting in that way.

What might a suitable candidate claim look like? Perhaps it is someone's end simply that people not drive cars from A to B. But it is not wrong merely to act in ways which preclude people from achieving particular ends: your end of being able to sit in your favourite seat on a public bus, for example, is denied by my act of sitting in it, but Kantians do not wish to claim this action is prohibited. It may perhaps be wrong to act in ways that conflict with others' ends when these ends arise directly out of their overall conception of the good: when their action just is one of their ends, rather than a means of achieving one of their ends. If someone had decided that what really matters in life is sitting in a particular seat on a bus, then as long as allowing him to do this does not conflict with your own ends, it

[14] The Kantian argument might be that all deception counts as denying the deceived party the *power* to set ends for themselves to some degree, even if, in the particular case, the absence of deception would not change which ends the person would choose.

would arguably be wrong not to offer him your seat: your benevolence shows you regard his capacity to set ends as having equal importance to your own, when you have no reasonable grounds for refusal. But, just as it is implausible that any such bus seat fanatics actually exist, it is implausible to suppose that my act of refraining from driving from A to B could count as intrinsically valuable to someone, such that it would count as benevolence to refrain, and I may have good reasons for travelling in this way.

The core cases in which the Formula of Humanity is applied are cases in which my acts do not simply frustrate the ends of others, but cases in which I deny people the *capacity* freely to set ends for themselves, by threats, by manipulation, by deceit, by violence. Whether any of these apply in the climate change case would seem to come down to whether I am reckless: whether I do not show sufficient respect for the humanity of others by acting in ways that limit their ability to act according to their conception of the good. Whether I as an individual do this through contribution to a collective harm must surely depend on whether I have an impact. If this is true, then, as a solution to the collective impact problem, the Formula of Humanity must be vulnerable to the same criticisms adduced against individualist consequentialist theories.

Kingdom of Ends Formula

I argued earlier that one may have most reason to perform some action, even if that action would be prohibited by a convention that a suitably placed legislator would have a duty to establish. In this context, it would be remiss not to address a final central Kantian concept, that of the Kingdom of Ends, understood as 'the systematic conjunction of rational and reasonable persons under common (moral) laws' (Rawls 2000 208). According to this formulation, we should regard ourselves as legislating laws for our shared moral commonwealth. What is distinctive about this formulation, on John Rawls's reading, is that it draws 'attention to the *mutual recognition* of the moral law in the *public* role of society's

moral culture' (Ibid. 209, emphasis added). This does point towards a social coordination function for the categorical imperative: it seems to mandate a kind of tendency towards optimism with respect to the capacity of our fellow moral agents to do what the moral law requires. This speaks to the idea that taking a "Kantian attitude" by taking a cooperative stance when one finds oneself in a coordination problem is in some sense self-justifying. As Jon Elster (1985) observed, when faced with collective action problems, we would all rationally want as many of the other participants as possible to be Kantians, in the sense of being disposed unilaterally to engage in cooperative behaviour, and this fact itself may seem to recommend a Kantian approach. Similarly, Marion Hourdeqin (2011) has argued that because collective action problems are more 'tractable' when participants have a tendency to be unilaterally cooperative, we should adopt a norm whereby cooperative behaviour is required of us as a default, irrespective of the behaviour of others (Hourdequin situates her view within a broadly virtue ethics-based approach, rather than Kantianism).

As Elster also points out, however, this attitude can in certain contexts be regarded as treading dangerously close to magical thinking. Consider the following piece of reasoning, which might be offered as a solution to the question of why one should vote, given one's vote makes no difference to the result (the so-called voter paradox):

I am a fairly typical member of my political reference group. If I vote, it is pretty likely that others will vote as well. Being like me, they will tend to act like me. Hence I shall indeed vote, to bring it about that others vote as well. (Elster 1985 144)

This is a kind of fallacious reasoning reminiscent of Newcomb's paradox (see Nozick 1969). Just because I am representative of my social group, it does not follow that there is a causal relationship between my behaviour and the group's behaviour. The analogy with the Newcomb problem is especially clear if you imagine everyone else already to have voted: the voter would then be attributing to himself the power to change the past, as well

as some mysterious power of social control. Is Hourdequin's reasoning structurally analogous to Elster's magical voter? Hourdequin should not wish to make a consequentialist argument: she should not wish to say that what gives the individual a reason to comply unilaterally is that he thereby makes full compliance *more likely*. This would be reasoning of precisely the kind non-consequentialist theorists are trying to avoid. Unfortunately for Hourdequin, there is little else that she could mean, without falling into the kind of irrationality Elster describes. It is not the case that, because in the optimal outcome I behave cooperatively, I ought to behave cooperatively in the actual world. Similarly, just because for Elster's voter, the outcome on which his party wins a decisive victory is one in which he and everyone like him votes, it does not follow that he ought to vote for that reason.[15]

A noted problem with Kantian "unilateralism" is that it may lead to outcomes that conflict with common sense moral principles about the avoidance of disaster. The doctrine of Unilateralism in international security could be a perfect example of this. Although the optimal outcome of international cooperation in nuclear disarmament would be a world in which the USA, along with everyone else, had no nuclear weapons, it does not follow that the world being as it is, the USA ought unilaterally to disarm itself of all nuclear weapons. To do so might not just damage the interests of the US, it could make the world as a whole more dangerous. In the worst-case scenario, it could precipitate a large-scale nuclear war (it is at least possible it could – the politics of nuclear disarmament are not relevant for

[15] Of course, unlike a voter in a secret ballot, it is not the case that unilaterally reducing one's carbon footprint cannot have a causal effect on the behaviour of others. One can influence the behaviour of others through one's example. As a consequentialist argument, this gets us only a little further. The social influence of certain individuals may perhaps be very great, but for people with no particular celebrity - which is to say the vast majority of us - it is equally possible that our influence is negligible. And at any rate, this form of argument only provides a reason for convincing others not to emit: if one is able to do this while continuing to emit oneself (in secret, for example), it would give one no reason to do otherwise.

present purposes). If it is true that Kantian reasoning is determinately at odds with common sense morality for at least some coordination problems, we might reasonably conclude that it cannot be trusted to give the right result with respect to any coordination problems.

That said, as Allen Wood argues, it may be methodologically misguided to dismiss Kantianism for its apparent failure to cohere with common sense morality – and this goes for many of arguments adduced against Kantian principles in the foregoing sections. Wood regards Kantian methodology as proceeding from 'a fundamental principle whose ground is independent of moral intuitions or Common Sense', and is rather 'an articulation of a basic value' – the value of rationality (Wood 2011 60). Particular duties then represent 'an interpretation of the normative principles applying that basic value under the conditions of human life' (Ibid.), and where the derivation of these duties appears to conflict with common sense morality, this should be resolved by reinterpreting the content of the duties in light of the fundamental principle, adding exceptions if necessary. This is to be contrasted against the method employed by Parfit, which also characterises much of contemporary moral philosophy, according to which moral principles are adduced and then refined by testing them against common sense moral intuition, until they seem to get the "right" answer in every case to which they are applied.

For a theorist like Wood, then, when Kantian principles apparently conflict with common sense morality – as applied, for example, to the individual's duty to contribute to public goods in circumstances of non-compliance - the way to proceed is not to consider those principles debunked, but to consider how they can be reinterpreted, guided by the fundamental principle that human rationality is an objective end and the source of all objective ends. If this is right, then the truly Kantian solution to the problem of collective harm may not be to find some principle that seems to guide individual action in the way that appears most appropriate, but to find some means of achieving the result which seems

obviously most consonant with the recognition of rational humanity's place at the centre of nature – the preservation of humanity's natural life support system. This is a very broad answer, and perhaps not a particularly helpful answer, but it is very difficult to say that it is the wrong answer. Thus Kantians may have resources to claim that their response to the problem of collective harm is not actually mistaken, on sophisticated readings, but for our purposes, their response remains unsatisfying.

Non-Consequentialist Individual Approaches II: Virtue Ethics

There seems to be a profound affinity between environmentalism and an ethic based around concepts of virtue. Some environmentalists have come to the conclusion that the dominant forms of ethical theory in western culture – namely utilitarianism and the liberal theory of individual rights – are partly to blame for the climate crisis. The thought, in its crudest form, is that these theories have led to a technocratic mindset, in which humans see their moral role as one of modifying the world around them, maximising the good (usually through the promotion of economic growth) and preventing rights violations (by coercing wrongdoers). They have lost, the thought goes, a more primordial concern for living well, for aspiring to be of good character - an approach to ethics that does not presume humanity's ability to control the world around it, but rather points to the value of traits like precaution, humility and mindfulness. Virtue theory has also been said to score over mainstream moral theory for the environmentalist because it eschews the perceived anthropocentrism of mainstream theory (see e.g .Hill 1983, Sandler 2016). For the virtue theorist, we should value the natural environment simply because this is what the virtuous person does. Consequentialists and deontologists, meanwhile, argue that we should protect nature only insofar as it promotes the good on some particular conception, or insofar as it serves persons as objective ends, leading, the thought goes, to a grudging and conditional conception of nature's value.[16]

Theorists who advocate an ethic that prioritises concepts of vice and virtue come to the view on the basis of various pieces of justificatory architecture. Some advocate a two-level view, according to which inculcating respect for certain virtues is regarded as justified

[16] Of course, this criticism only applies to standard versions of consequentialism and deontology, it does not apply to them by definition. One might defend a version of consequentialism according to which non-human natural phenomena are considered a final value (see Hiller 2013), or a version of deontology for which non-human natural phenomena are rights-bearing objective ends (see Regan 1983; Wood 1998).

because it serves some other ethical value, such as producing optimal states of affairs. Others come to the view via Aristotelian naturalism, where ethical value is viewed as just one part of the general concept of the good, such that to be an ethically good human being is just to be a good example of that kind. Some come to virtue theory from a social constructivist metaethics, whereby the virtues are those qualities that are regarded as admirable, and qualities that are regarded as admirable are those that tend to promote a well-integrated society. Many virtue theorists would want to dissolve the problem of collective harm by treating it as an outgrowth of an ethical system in need of radical revision. Chapter 5 is an analysis of the call for radical revisionism itself and how seriously we can take it in the context of the climate emergency. Here, I will confine myself to outlining and evaluating the work of virtue theorists who have responded to the collective harm problem more directly.

Ronald Sandler (2010) argues that a virtue ethics approach can be recommended precisely because it offers an adequate response to the collective harm problem - or as he calls it 'the problem of inconsequentialsm' - where approaches including Kantianism and Utilitarianism fail, on his account. He argues that an adequate ethical response to the problem of inconsequentialism would be one which explained why individuals should 'make the effort' or 'take on costs' to address environmental group-caused harms. Utilitarianism, it is claimed, fails by this measure, because it implies one should only take on costs when by doing so one can produce greater benefits. As it is not the case that one determinately produces benefits when reducing one's emissions, taking on costs cannot be justified. In response to utilitarianism's failings, he claims that an ethical theory that can adequately respond to the problem of inconsequentialism will have features utilitarianism lacks – it will not evaluate discrete actions on the basis of outcomes, but rather it will evaluate patterns of behaviour throughout someone's life. That theory is virtue ethics. Dale

Jamieson similarly argues that the right moral approach to deal with group-caused harms is one that instantiates 'non-contingency', meaning the evaluation of behaviour should not be contingent upon what others do (Jamieson 2007). He points to virtue ethics as such an approach. Marion Hourdequin observes the 'communicative moral value' of adopting a unilaterally cooperative stance in collective action problems (Hourdequin 2011). This, I take it, is not a consequentialist appeal the beneficial effects of setting a good example, rather, it is related to the virtue of 'integrity' to which she appeals in an earlier paper (Hourdequin 2010). One shows integrity, conceived as a consistent commitment to one's values, the thought seems to be, if one reduces one's emissions without waiting to see whether others will do the same.

These approaches face a similar objection to the one faced by Kantian "unilateralism", as we saw in the previous section. Hourdequin, Jamieson and Sandler all seem to start from the premise that the disposition unilaterally to cooperate is the ethically appropriate disposition to have in situations of group-caused harm, because if everyone had this disposition, the group-caused harm would not arise. This neglects to account for situations in which unilaterally "cooperative" behaviour fails to contribute to any harm prevention, because too few others follow suit. There could moral costs, as well as economic costs, to accepting burdens without the expectation of benefit. Imagine, for example, an overzealous environmentalist whose desire to reduce her individual carbon footprint led to her to neglect her own family. It is unclear, in other words, why these theorists are so sure that the virtuous person *is* unilaterally cooperative. Why not regard the person who decides to drastically reduce her emissions, even though there is no expectation of anyone else doing the same, as viciously self-destructive, as failing to show the proper concern for her own flourishing?

Virtue ethicists have a ready-made response to these worries: the dispositions that make up the virtuous character are to be judged in the round: someone who reduced her carbon footprint to such an extent that she inhibited her own ability to flourish would not be regarded as a virtuous character. This response introduces the vexed question of from where virtues are to be drawn, of how we are to know what the virtuous person looks like. The idea that the system of virtues should form a mutually supportive whole, making it impossible that traits that made one virtuous in one respect could make one vicious in other, dates back to the virtue theorists of antiquity. Alastair MacIntyre, however, a key figure in the rival of virtue ethics, is among those who explicitly deny the doctrine of the Unity of Virtue, viewing this borrowing from Plato in Aristotle's work as siting uncomfortably with important aspects of Aristotle's theory. Specifically, MacIntyre holds that the central place Aristotle assigned to *phronesis*, or practical wisdom, reflects a recognition that a theory of the virtues should be highly responsive to circumstance, and that cleaving to theoretical harmony unduly hinders a theory on this measure (MacIntyre 1981 143). Though this view is controversial, we must acknowledge at least that the Unity of Virtue is contestable and should not be treated as an article of faith.

Whatever our view on the Unity of Virtue, the question of what justifies given traits as virtuous remains problematic for the application of virtue ethics to group-caused harm. If, like MacIntyre, we view the recognition of virtue as inherent in social and political practice, and heavily embedded in tradition, then given that participating in the fossil fuel economy is generally regarded as normal and unobjectionable behaviour in societies as we presently find them, it would be hard to see why a fossil fuel emitter would have to be regarded as a

deficient character.[17] On some conceptions of virtue ethics at least, the view takes social practice as its guide to virtue, making it in a certain sense inherently conservative.

Perhaps, on the other hand, we have a naturalist conception of virtue, where what makes some trait virtuous is whether it supports its bearer's natural functioning. The judgement that some human animal lacks virtue is then analogous to the judgment that some non-human animal is a deficient example of its species, like a wolf that refuses to hunt with the pack, or a sparrow that fails to build a nest (see Hursthouse 1999 192; Foot 1995). If this is true, then in distinguishing virtue from vice we could be making a discovery. If it is newly established that certain tendencies make us bad examples of humanity, then we have new objective grounds to condemn them, even if those tendencies are widely accepted in human societies as they currently exist. The discovery that certain carbon-intensive activities currently regarded as normal, such as driving cars and taking flights on aeroplanes, are contributing to the degradation of nature's ability to support human life, it could be argued, is just the sort of discovery needed to establish these activities as vicious.

The problem with this form of argument is that the kinds of traits that (on the most plausible available conceptions) constitute an individual's natural functions are not necessarily what is best for *other* individuals of its kind. A supremely successful solitary animal might outcompete other members of its species for food or mates, or a successful group of animals might outcompete other groups. For example, a pride of lions might dominate a territory so effectively that they prevented any itinerant males in the area from establishing new prides. Success at an individual level might even lead to species collapse

[17] This argument presupposes that the traditions and practices according to which high-consumption lifestyles are considered normal cannot be viewed as in some way moribund from an internal perspective. MacIntyre's project was fundamentally critical of prevailing moral attitudes, he would certainly be sympathetic to the view that some aspects of the ethics of consumer capitalism were detached from the conception of virtue embedded in healthy forms of life. Nevertheless, it is difficult to see how a virtue ethics grounded in traditions and social practices has the resources to condemn fossil fuel emitting behaviours *as such*.

– take lemmings, for example, which fulfil their natural function well by producing large numbers of offspring, only for the next generation to die on mass due to a lack of resources.

Sociality and rationality are typically regarded as the central functions of humans, rather than mere survival. But this does not mean an analogous story cannot not be told for humans, *mutatis mutandis*, as for lemmings or lions. It is sadly not so difficult to foresee a radically climate-damaged future, in which a now much smaller global population, the descendants of the very richest people alive today, live insulated from the ravages of climate change in some relatively unspoiled corner of the world, while the rest have been consigned to war, famine, pestilence and death. Why not regard this exalted class, who, we may imagine, have perhaps attained new heights of social and rational sophistication, as the standard-bearers of human flourishing, and the people scrabbling in the dust beyond the walls as failed examples of their kind? The appeal to natural teleology, in short, is cogent only as long as it defines the *telos* of humankind, and the nature of the threat posed by putatively vicious behaviour, in a highly tendentious manner.

The virtue ethicist might retort that it is not simply self-serving to characterise the *telos* of humankind in a way that rules out the apocalyptic vision of the previous paragraph as a desirable form of life, because virtue ethics is undergirded by a particular vision of humankind's relationship with nature. Virtue ethicists evaluate goodness in terms of natural teleology not only at the individual and species level, but also at the level of broader natural categories like ecosystems. Individual organisms, humans included, are evaluated not just by the quality of the support their parts provide to the functioning of the whole organism, but also as parts themselves, according to the quality of the support they provide to the natural systems in which they are entwined. The thought then would be that even if the fossil fuel economy allows humans to carry out their individual natural functions to a high degree of success - using their rational capacities to master nature, free themselves from

want, and enjoy all the social pleasures of friendship and love within a well-maintained social order - they would still be failing as good specimens of their kind if participation in that economy damaged the natural systems that allowed other beings to fulfil their natural purposes.

This kind of reasoning may well rescue the environmental virtue ethicist from the charge that she performs a rabbit-and-hat trick by producing her desired conception of environmental virtue from a pre-loaded description of human nature.[18] However, the view faces a more fundamental problem specific to its application to the present subject: its focus on characters rather than acts obscures rather than elucidates the problem of collective harm. Say temperance is a virtue, and excessive consumption a vice. Why should one conclude that any particular act of consumption is viciously intemperate? Indeed, as a constitutive matter, one may not: a virtue is a set of dispositions that persist over time, a character trait, and therefore no particular act is virtuous or vicious, as vice and virtue are properties of characters and not of acts. Virtue ethicists including Hursthouse have loosely defined the virtue theoretic conception of right action through the claim that an action is right if it is what the perfectly virtuous person would do in the circumstances (Hursthouse 1999 25). Collective harms, however, can arise from the aggregate effects of one-off acts, if the group of individuals perpetrating such acts is sufficiently large. The perfectly virtuous person could, in theory, perform one such isolated action, and still qualify as perfectly

[18] This is not to say the point cannot be resisted. We might argue that the problem of lions and lemmings re-emerges at an ecosystem level. If the good as it pertains to organisms is (at least partly) constituted by fitness for survival, then by analogy one might think the good as it pertains to ecosystems is constituted by the tendency to maintain an equilibrium state over a long period of time. The most complex ecosystem is not necessarily the most stable. In fact, mathematical ecologists believe the reverse is true: ecosystems with fewer elements are more resilient ("Will a Large Complex System be Stable?", May 1972). If this is right, then extinction events might actually be good, from an ecosystem perspective, insofar as they bring the system down to a more stable state. In this sense, just as the good of the pride is not necessarily the good of all individual lions, the good of the ecosystem is not necessarily the good of all species. It therefore remains somewhat tendentious to claim that a human that damaged the ability of an ecosystem to support certain other species would by that token be a poor example of its kind.

virtuous, as long as she was responding to good reasons. An ethics of virtue, then, is structurally unsuited to responding to collective harm problems, except by observing that a world in which everyone were a paragon of virtue would be a world in which fewer collective harms occurred. While those sympathetic to virtue theory may regard it as dissolving the problem, others may regard it as simply ignoring it.

Furthermore, Hursthouse argues that virtue ethics is not 'committed in advance to our living well being a realisable state of affairs regardless of how we, or how many of us, have lived up until now' (Hursthouse 2007 170). It may be that, because we live in fossil-fuel dependent economies, and because so much greenhouse gas has already entered the atmosphere, no way of living in accordance with environmental virtue remains open to us. The putative virtue of 'being rightly oriented with respect to nature' (Ibid.), for example, conjured up by Hursthouse among other theorists, may simply be impossible even to approach, and therefore unavailable even as a regulative ideal or a best-case scenario. If one attempts to leave society altogether in order to achieve proper orientation towards nature, Hursthouse points out, one 'will have cut [oneself] off from the exercise of most other virtues' (Ibid.). If this is right, then virtue ethics converges with the bullet-biting response: it concedes that although the collective harm of climate change is clearly a grave problem, it is not specifically a problem for ethics.

Coordinated Group Approaches

If the occurrence of collective harm involves a paradox, consisting in the idea that it appears someone or something should bear outcome responsibility for the harm, and yet there are no individual persons to whom responsibility can reasonably be attributed, then perhaps the solution is that responsibility should be attributed not to individuals, but to groups taken as a whole. The idea of collective responsibility has a chequered history, with some liberals being sceptical of the concept, or indeed downright morally outraged by it, on the grounds that it fails to 'take seriously the plurality and distinctness of individuals', to apply Rawls's formulation (Rawls 1971 29). If the attribution of collective responsibility is anything other than a convenient shorthand for attributing individual responsibility to each member of the group, this brand of liberal fears we run the risk of scapegoating the innocent and absolving the guilty.

There are two sorts of groups for which a longstanding tradition exists of regarding them as the sort of entities capable of bearing responsibility: corporations and states (or perhaps, specifically nation-states). The idea that organisations can have legal or moral personality, and thus bear responsibility, goes back at least as far as the reign of Pope Innocent IV, who introduced the concept of *persona ficta*, partly as a means for monks to engage in trade without breaking their vows of poverty: monasteries established the legal convention that the monastery itself was the owner of property, rather than any of the individual monks. Once one has a concept of corporate property, it is a short step to the idea of corporate liability: no individual monk would necessarily be on the hook to make good on any contract the monastery had concluded, but the monastery itself would be. States, similarly, are another example of a collective entity standardly viewed as having personality (Crawford 2006). In Commonwealth jurisdictions, for instance, there exists the legal entity called the Crown, which is not identical with the monarch of the day as an individual, but which

represents the functions of the state, and which inheres in the monarch of the day during his or her reign.

Both monasteries and states are suitable for bearing personality because they have agency – even if their agency is in practice coterminous with the agency of a particular individual, the abbot of the monastery in question, or the monarch of the day (whose authority is delegated to the various branches of government). The principle of "ought implies can" makes agency a necessary condition for an entity to be capable of bearing a duty. As our interest in responsibility is usually closely related to our interest in the assignment of remedial duties, agency would therefore seem to be a necessary condition of group responsibility. There are a number of theories of collective agency, which give rival specifications of the conditions a group of individuals must meet in order to be capable of group agency. Some believe that for a collective entity to bear accountability, it must be a group agent. Standardly, this is said to involve some condition regarding the ability to make decisions at the group level, and a condition specifying the entity must be capable of behaving rationally over time (French 1984, Petit and List 2011). Others believe that groups can bear responsibility which are not group agents proper, but which exercise agency together in an *ad hoc* way, through shared intentions (Gilbert 2013), or individual participatory intentions (Kutz 2000).

The problem for our purposes is that none of these accounts straightforwardly applies to climate change as an instance of the problem of collective harm. Although a large proportion of global greenhouse gas emissions can be attributed to the activities of corporate entities (see Introduction), there is significant proportion that is arguably best viewed as being produced by a large group of individuals acting independently of one another, widely dispersed across time and space, which therefore cannot be regarded as partaking in any of the structures of coordination needed to establish a group agent or

collective agency more generally. In the next section, as well as the following chapters (see Chapter 4; Chapter 5), we will examine whether there are minimal forms of group coordination which give rise to reasons for individuals to regard themselves as responsible for what the group does, or as acquiring special duties in relation to what the group does. In what follows, we will briefly consider the extent and limits of state responsibility as a response to the collective harm problem.

Sceptics of individual responsibility for contributions to greenhouse gas emissions including Walter Sinnott-Armstrong point to the role of the state as the most appropriate agent for bearing responsibility for climate change. Appeal to state agency cannot provide a general solution to the collective harm problem, as collective harms could in principle occur in situations in which no appropriate state actors exist. Moreover, for the appeal to state agency to constitute even the right kind of response to the collective harm problem, specifically with respect to the responsibility gap intuition, it must be possible to assign outcome responsibility to states. For Sinnott-Armstrong, this is apparently not the claim he is making. Governments, he says, have obligations in relation to climate change 'because they can make a difference' (Sinnott-Armstrong 2005 312). John Broome, similarly, argues that the primary duties of governments with respect to climate change are duties of goodness rather than duties of justice (Broome 2012 97), which is to say states should be concerned about climate change primarily because they have the power to make the world significantly better, and therefore the duty to use that power, rather than because they should be concerned about repairing past wrongs or avoiding future ones.

Stephen Gardiner argues that 'political institutions and their leaders are said to be legitimate because, and to the extent that, citizens delegate their own responsibilities and powers to them' and therefore that 'the most direct responsibility for the current failure of climate policy falls on recent leaders and current institutions' (Gardiner 2011 53). For Gardiner,

then, citizens - or the group of individuals that happens to be affiliated with a given state - are the real bearers of responsibility for climate change, and states are their delegates in discharging that responsibility. If states fail, the responsibility is said to fall back upon the group of individual citizens, who then have a new duty to found institutions capable of discharging their responsibility. This is why Gardiner has more recently advocated a new supranational institution, directly answerable to world citizens, given the failure of states to agree and implement significant policy change via the United Nations Framework Convention on Climate Change (UNFCCC) (Gardiner 2014, 2017). As with Broome and Sinnott-Armstrong, then, Gardiner's approach does not attribute outcome responsibility to states for the harms of climate change. Rather, it attributes to states a moral failure in their fiduciary duty to their citizens.

Though states are regarded as having legal personality, in international law at least, their liability has historically been viewed as being limited to liability for violating norms of legitimate state conduct. It is not until relatively recently that a tradition began to be established whereby states could be regarded as responsible (liable) for harms perpetrated by actors on their territory against the citizens of other states, when such acts were lawful under domestic and international law. When such principles were codified, they simply imposed a duty of due diligence upon states to regulate activity that carried a risk of transboundary harm; they did not actually impose liability for the harm itself on the state (see Tanzi 2013). Successive agreements under the UNFCCC can be seen as continuing this tradition: states' responsibilities for loss and damage due to climate change have been limited to '[e]nhancing knowledge...strengthening dialogue, coordination, coherence and synergies among relevant stakeholders' and 'finance, technology and capacity building' (UNFCCC/CP/2012/8/Add.1 3/CP.18 para.5), with powerful state parties declining to

acknowledge a compensation duty of the kind that would be implied if states were regarded as bearers of outcome responsibility.

There is, however, a significant tradition in climate ethics attributing responsibility to states on the basis of historic responsibility. An overlapping tradition defends the "polluter pays principle" (PPP). A distinction can be drawn between a historic responsibility principle and a polluter pays principle, insofar as historic responsibility is backward-looking, attributing remedial responsibility on the basis of total historic contributions to GHG emissions up until the present, whereas the PPP is forward-looking, the claim being that polluters today should have to pay the true social cost of their emissions, meaning it is often cited in relation to carbon pricing (see Shue 1999). As applied to states, however, these principles might be thought to express the same intuition. Henry Shue argues that the best justification for these intuitions is a hybrid principle, which combines considerations of causation with considerations of benefit. It is both unfair and perverse, the thought goes, that anyone should be permitted privately to enjoy the benefits of carbon-intensive development while imposing the cost of doing so on the world as a whole. States, then, are the appropriate locus of responsibility because 'in an international system built around sovereign states most assets in fact simply continue in whichever state they are accumulated' (Shue 2015 22): because the benefits of the fossil fuel economy have been passed down through the generations within particular states, allowing subsequent generations to start from a higher baseline of development, particular states should today be the bearers of responsibility for historic emissions.

Another tradition, of which David Miller is a key exponent, holds that the locus of responsibility for climate change is neither states nor corporations, but nations or peoples. Miller maintains this view despite the fact such groups may lack mechanisms for making group decisions or forming group intentions. *States* characteristically have these capacities,

but nations are only contingently affiliated with state institutions, and Miller does not regard affiliation with such institutions as a necessary condition of collective responsibility. He offers two 'models' of collective responsibility, which he thinks both support the idea of national responsibility in their respective ways, although they are separable. These are the 'like-minded group' model and the 'cooperative practice' model. On the 'like-minded group' model, a group bears responsibility if its members share 'aims and outlooks in common' and 'recognize their like-mindedness' so that 'when individuals act they do so in light of the support they are receiving from other members of the group'. On the 'cooperative practice' model, a group can bear responsibility if its members are 'beneficiaries' of a common practice, where that practice involves its participants being treated fairly (Miller 2007 117). Nations, then, are said to be both like-minded groups, and cooperative practices, and for that reason, are to be regarded as the bearers of responsibility for climate change.

Neither invoking national responsibility or state responsibility solves the generalised version of the collective harm problem: not all groups that produce collective harms will be nations or states, therefore invoking national or state responsibility does not resolve the paradoxical absence of outcome responsibility, at least in all cases. However, national or state responsibility would go some way towards responding to the "responsibility gap" intuition in the case of global climate change. If by assigning responsibility to such groups we can assign remedial responsibility for the total amount of harm that will be caused by GHGs emitted up until the present day, we will at least dispense with the apprehension that some of the harms of climate change are in that sense morally unaccounted for.

Both the nationalist and the statist approaches are vulnerable to a common criticism. It is not clear that it is correct to identify either the group of beneficiaries of carbon-intensive development, or the group of people who share a like-minded commitment to the fossil

fuel economy, with either states or nations. As has been stressed by authors in the cosmopolitan tradition in relation to global justice, the spoils of economic activity do not respect state boundaries. Resources produced in one country fuel production in another, which are then purchased by end-consumers in a third. Though it is clear that certain regions have benefited disproportionately from historic carbon-intensive development (chiefly North America and Europe), those benefits do not neatly fall along state lines in proportion to historic emissions.

In the middle of the 19th century, the majority of the world's cotton was grown in the United States and shipped to mills in Britain (most powered by coal) where it was processed into garments which were then sold to consumers all over the world. The benefits of fossil fuel consumption in Britain, then, accrued not only to British mill owners, but also to plantation owners in the United States, import-export companies under various flags, and garment consumers across the world. Arguably, then, at that time, although the carbon emissions of the UK were proportionally much higher than those of many other countries, British carbon emissions fuelled growth and increased the standard of living far beyond the UK's borders. A similar argument is sometimes made with respect to China in the present day: although China has the highest annual GHG emissions of any country in the world, it is not necessarily true that the benefits of those emissions are felt exclusively by Chinese citizens. Chinese industry has supported the global consumption of cheap consumer goods, which (by some metrics at least) have increased standards of living in many other countries. Thus it arguably unfair to use total historic emissions produced on the territory controlled by a particular state as a yardstick for that state's responsibility in the present day, on the grounds that the benefits also stayed within the territory.

Nations, similarly, cannot be easily counted as the locus of beneficial 'cooperative practices' that inherently involve the fossil fuel economy. Miller is not wrong to suggest that nations

can be regarded as cooperative practices in the sense he intends the term, namely that via our membership of nations we gain benefits which we regard as engendering claims of fairness. A French person, for example, would tend to regard the continued existence of the French nation as valuable. She would tend to object to policies she perceived as changing important aspects of French national culture - secularity, for instance - implying she regards herself as having some sort of claim over them. Nor, of course, is Miller wrong to suggest that members of nations benefit from the fossil fuel economy. But what is more problematic is the suggestion that people benefit from fossil fuels *as* members of nations, or that the fossil economy can be regarded as a beneficial aspect of national culture.

Clearly, some aspects of fossil fuel consumption are tied to some aspects of nationhood. Coal mining, for instance, would perhaps at one stage have been regarded as an important aspect of the national culture of Wales (and may still be regarded as such by some). Car ownership and the consumption of cheap fuel may be considered an aspect of American national culture by many (although this is by no means universal: denizens of coastal cities, for instance, might be more likely to deny it). But such examples are few and far between, and those that are acknowledged can be controversial, even divisive. Miller, it seems, wants to make a slightly broader point, commenting, '[w]e say, for instance, that Germans are hard-working, meaning that the way individual German workers behave reflects a shared norm of industriousness that forms part of the public culture of Germany' (Miller 2007 126-127). Similarly, perhaps, Americans are viewed as entrepreneurial, the Japanese as very dedicated to their jobs, Brits as natural traders ('a nation of shopkeepers'). Certain stereotypes about national character are seen as entwined with the capitalist mode of production, or with high-consumption lifestyles more generally, and insofar as these forms of life are dependent on fossil fuels, perhaps Miller can assert the connection between nationality and the benefits of the carbon economy that he needs.

The problem with this appeal to national stereotypes is it cuts both ways. For ascriptions of national responsibility to be valid, Miller writes, 'it is clearly crucial to establish that their collective actions are a genuine embodiment of the shared beliefs and values that go to make up the national culture' (Ibid.). This makes attributions of national responsibility contingent on cultural phenomena that are by their nature highly dependent upon interpretation, upon the manner in which they are descriptively contextualised. What are we to say about outdoorsy Canadians, happy-go-lucky Irish, laid-back Italians? In the case of many nations, it is much more difficult to produce a stereotypical narrative that links their productive activities in relation to fossil fuel consumption to their national culture. Yet it seems wrong that a large group of individuals organised as a nation should be able to evade responsibility for collective harm simply because they regard their carbon-intensive development as having been somehow reluctant, or tangential to their shared values.

Even Miller's own examples bring out the difficulty in linking high-consumption culture with nationhood – he refers, for instance, to 'the pattern of family relations in a particular country, and the number of children who are on average produced' corresponding to 'the religious or other cultural values of the nation in question' (Ibid.), as a potential source of outcomes for which the nation could be held accountable. But Roman Catholicism, for example – a religious belief system that correlates with higher birth rates – is a transboundary cultural practice. If collective responsibility is to be grounded in the claim a certain outcome is 'the genuine embodiment' of a particular cultural practice, why is collective responsibility not to be vested in the group of adherents to that specific practice, rather than to a national culture? National culture can arguably be regarded as consisting in a whole cluster of partially overlapping practices, each of which might transcend borders and might count only a minority of the citizens of a particular nation-state as participants. Miller's account of national responsibility invokes the idea that people should be

accountable for harmful cultural behaviours, making it a kind of indirect causal responsibility. The role of nationality itself, then, becomes somewhat redundant, as it seems to be participation in specific cultural behaviours that is doing the normative work (see Chapter 4, Chapter 5).

None of these arguments should be taken as counting against the attribution of *remedial* responsibility to states. As Shue is keen to stress, it may be that a constellation of mutually supportive considerations, from ability-based arguments, to beneficiary-pays reasoning, to the considerations of either direct or indirect liability for damages, all converge on the conclusion that states are accountable for GHG emissions. It is not my project to undermine such attributions of accountability. I am certainly sympathetic to Shue's desire to find theoretical shortcuts when these give us the weapons we need to oppose those who want to obfuscate the normative demand for an adequate response to the climate crisis, especially on the part of the most powerful state actors. From a theoretical perspective, though, it is important to have a clear account of whom we should regard as the primary bearers of outcome responsibility for climate change, even if in practice conflicting accounts will have similar policy implications. There is also political value in getting it right. A conceptually clear account of the ground of responsibility is a bulwark against what Stephen Gardiner has called 'moral corruption': the tendency to use theoretical obscurity as an excuse for inaction (Gardiner 2011 302).

Uncoordinated Group Approaches

As we have seen, we are primarily interested in assignments of outcome responsibility because we are interested in assignments of remedial responsibility. Uncoordinated groups – mere collections of individuals without structures of shared intention, assigned roles for particular members, or procedures for decision making, cannot act, and therefore cannot bear remedial responsibility, at least not in any direct sense. There is, however, a tradition going back at least as far as work by both Joel Feinberg and Virginia Held in the late 1960s and early 1970s, which argues that uncoordinated groups can in fact bear duties. If this is right, then it is becomes less controversial to claim that uncoordinated groups, such as the group of emitters, can indeed be regarded as bearers of outcome responsibility. Elizabeth Cripps (2013) can be read as defending a version of this claim.

Held's argument for the claim that an uncoordinated group, or a 'random collection' of individuals, as she called it, could bear responsibility, appealed to a case of seven strangers in an underground train carriage (Held 1970). The second smallest of these strangers begins strangling the smallest, and, after a protracted struggle, kills him. The other five would together have been able to subdue the attacker easily if they had worked together, and some collection of fewer than five would likely have also been sufficient, although no one of them alone would have been able to subdue the attacker without great danger to themself. Held observes that with respect to a case such as this, "the group is morally responsible for the victim's death" is a natural judgement. Feinberg presented a similar case, considering a group of train passengers being robbed by the famous outlaw Jesse James. A sufficiently large subset of the passengers working together would have been able to subdue James, but no individual passenger could be considered culpable for failing to make an attempt on him, as to do so would have been an act of exceptional heroism. Feinberg argues that it makes sense to regard the group of passengers taken together as

culpable for failing to subdue James, because a kind of fault was indeed involved, although not the fault of any individual – there was 'a flaw in the way the group of passengers was organised' (Feinberg 1968 687).

The problem with applying reasoning of this kind to the climate change case is it is not clear what follows from attributions of group responsibility of this kind. As Feinberg noted, group *liability* might not follow from group responsibility in these cases. Without an attribution of liability, there are no clear implications with respect to the allocation of costs. It would not be right, for instance, to expect the elderly lady at the back of the carriage who managed to keep herself hidden during the robbery to participate in a scheme to distribute the burden of compensation for the victims' losses. Held and Feinberg can be read as pointing out that it is reasonable to locate fault at the group level; what is less straightforward is the question of what, if anything, such fault implies with respect to remedial responsibility.

More recently, Elizabeth Cripps has defended an account of the collective responsibility of the group of polluters with respect to climate change that can be seen as building on the kind of intuition raised by Held and Feinberg, while doing more to flesh out the problematic relationship between collective responsibility and remedial duties. She claims that '[a] number of individuals who do not yet constitute a collectivity (either formally, with an acknowledged decision-making structure, or informally, with some vaguely defined common interest or goal) can be held collectively morally responsible for serious harm (fundamental interest deprivation) which has been caused by the predictable aggregation of avoidable individual actions' (Cripps 2013 68-69). She calls this a 'weakly collective responsibility', as opposed to the strongly collective responsibility that would be held by a corporate entity or other coordinated group, with group decision-making procedures or shared intention. It is 'weakly collective', in the sense that 'the result (harm) could not have occurred were not those individuals situated, in relation to one another, in such a way that

their pursuit of individual goals would have a certain predictable aggregative impact', and therefore that harm arises from 'the way the individuals are grouped' (Ibid.).

Cripps, then, makes use of a spatial metaphor to explain her conception of weakly collective responsibility: the group of polluters, though uncoordinated, bears collective responsibility in the sense that they are 'situated' in a certain way with respect to one another. In the examples given by Feinberg and Held, this was not a metaphor: the train passengers really were joined together by their similar proximity to a particular event. In the case of the group of polluters, however, the group is joined together by the predictable relationship between the type of act they perform and a certain outcome.

The claim, then, is that this relationship also has specific implications in terms of remedial responsibility. Weakly collective responsibility implies an individual duty to promote the formation of a collective capable of bearing strong collective responsibility and of discharging the remedial responsibility it implies, and a duty to play one's assigned part within that collective, once formed. Why should we think it has this implication? Cripps thinks an individual becomes a joint bearer of collective responsibility if she performs an action which contributes to a collective harm, providing she exceeds a level of contribution such that, were everyone to contribute to that level, 'there would be no harm' (Cripps 2013 73).

The condition that individuals should be held responsible if their contribution exceeds the level at which, were everyone to contribute to that level, there would be no harm is intuitively plausible, but would benefit from further justificatory support. In that sense, it is similar to Parfit's principle that an act 'may be wrong if it is part of a set of acts that together harm other people' (Parfit 1984 70), which Christopher Kutz, correctly in my view, called 'pure fiat' (Kutz 2002 480). Parfit's principle fails to explain *why* being part of a group that together causes harm should itself be considered wrong, or as incurring considerations of

remedial responsibility. Although Cripps offers such an explanation, more work is arguably needed to fill in this picture.

Cripps points to Feinberg's principle that for some condition to be considered a cause of harm, it must constitute a 'deviation...from the normal course of things' (Feinberg 1970 202, cited in Cripps 2013 73). The thought is apparently that contributions above the level below which, if everyone contributed to that level, there would be no harm, are exactly such abnormal occurrences. But this consideration is arguably not relevant to the question of the connection between contribution and weakly collective duty. The group of agents whose contributions to GHG emissions together cause some harm is – uncontroversially – the cause of that harm. This is the only fact Feinberg's principle picks out. If Cripps were to assert that any *individual's* contribution is abnormal, and therefore a cause, she would beg the question, as the collective harm problem is characterised precisely by the claim that no individual can be considered the cause of harm. These two interpretations (abnormality at a group level and abnormality at an individual level) exhaust the ways in which Feinberg's principle could be applied in the way Cripps wants. Neither serves to connect contribution to weakly collective duty.

It is also worth noting that Cripps's principle may cast the net of responsibility very widely. Even since the time Cripps was writing, it has become even more indisputably clear that if global emissions could somehow safely be reduced to zero overnight, the impacts of climate change, many of which are already ongoing, would continue to be felt for decades to come. Thus, if the term 'everyone', as it appears in Cripps's principle is read as "everyone alive at present", the principle leaves contributors to GHG emissions on the hook for the harms of climate change however small their contribution. This does not necessarily count against the view, but it does make for a very large and undifferentiated group of people who bear climate responsibility.

New Directions

Cripps has provided a model for how a middle ground between individual and collective responsibility for climate change could give us grounds to attribute remedial duties to individuals, through the idea of 'weakly collective duty'. This kind of approach represents the best chance of grounding individual remedial responsibility for climate change on the basis of considerations of outcome responsibility. In Chapter 4, I will argue that we can get close to justifying the attribution of outcome responsibility to individuals for climate change by describing a minimal form of shared agency, and begin to set out what that form of agency would look like. In the next chapter, we will return to the question of individualist approaches to the problem: some recent accounts have presented an initially compelling case for the claim that individuals have a duty to mitigate arising from duties to other individuals, in particular the duty not to harm. A close examination of why these arguments are unsuccessful will prepare the ground for Chapter 4's group-agency approach.

3. Individual Direct Duties: Three Case Studies

This chapter examines whether individuals have direct duties to avoid contributions to GHG emissions. By direct duties, I mean duties one person holds to another. For example, Broome's argument, of which we gave a brief analysis in Chapter 2, is an argument based on the duty one person has to another not to harm that other, and therefore an argument for a direct duty to avoid contributions to GHG emissions. I will argue that the best candidate arguments for individual direct duties to avoid contribution to GHG emissions fail, and therefore that it is highly likely that all such arguments fail. If other arguments for individual accountability for contributions to GHG emissions are available, they are to be strongly preferred. In other words, the argument in this chapter is entirely negative. It is, however, intended to prefigure the argument that if we have strong reasons to avoid individual contributions to GHG emissions, they are not reasons of individual direct duty, but reasons that arise out of our membership of groups.

As noted in the introduction, there are several possible framings of the problem of collective harm. One framing consists in the claim that the individual lacks a *reason* to refrain from contribution to the collective harm. Clearly, if one lacks a reason to refrain from contribution, then *a fortiori* one lacks a duty. Justifying the claim that that the individual has a reason to refrain from contribution to collective harm can be seen as a first step towards justifying a duty. This chapter will therefore first examine an important recent argument in favour of such a reason, due to Julia Nefsky. As we shall see, this argument justifies at best an extremely weak reason which is easily outweighed by other considerations. This demonstrates that arguments for individual direct *duties* to refrain from contribution to GHG emissions are not being sown in fertile ground. Next, we

examine Lawford-Smith's view, which appeals to the notion of thresholds effects to justify the idea the individual is directly responsible for harm. Rather than using thresholds to support the claim that individuals have an obligation to avoid GHG emissions because of the *expectation* of harm that would be attached to emission-causing actions (see Kagan 2011), she argues that individuals can be regarded as wholly responsibly for directly causing harm, because they are a necessary part of a group that jointly causes harm. We shall argue that this view has absurd implications. Finally, we will look at Broome's view, drawing on recent work that can be regarded as a new account. Broome's principal argument is that individuals have a duty to refrain from contribution to GHG emissions because doing so produces an expectation of harm, and this itself is wrong. As we shall see, this claim lacks sufficient support.

Reasons to Avoid Contribution to Collective Harm

Julia Nefsky proposes a solution to the problem of collective harm intended to show that individuals have good reasons to refrain from contributing to collective harms. Her case begins by observing that the idea individuals are *causally involved* in collective impact can be considered a given feature the problem, rather than a something to be demonstrated in offering a solution. If we did not know that individuals were, for example, part of the cause of climate change, there would be no special moral problem surrounding the nature of individuals' obligations; the challenges involved would be purely political. The problem, she holds, rather fundamentally consists in the fact we individuals can apparently find no reason not act in a way that makes one part of the cause of a harm, or no reason to act in a way that makes one part of the cause of benefit, when doing so would make no difference to the degree of impact that eventually obtains. What we are looking for, she holds, is a reason to act or to refrain from acting in ways that contribute to collective impact.

Take by way of example a case now familiar in the literature, originally due to Jonathan Glover, which has been discussed extensively by Derek Parfit and his commentators – the Drops of Water case.

Parfit describes the case as follows:

> *A large number of wounded men lie out in the desert, suffering from intense thirst. We are an equally large number of altruists, each of whom has a pint of water. We could pour these pints into a water-cart. This would be driven into the desert, and our water would be shared equally between all these many wounded men. By adding his pint, each of us would enable each wounded man to drink slightly more water—perhaps only an extra drop. Even to a very thirsty man, each of these extra drops would be a very small benefit. The effect on each man might even be imperceptible.* (Parfit 1984 76)

There are various ways of expressing the problem, but as Nefsky understands it, the central point is to recognise, first, that the valuable outcome in which we are interested is relieving

the thirst of the stranded group. Second, given that any individual gives only a single drop to any other individual, and given that one drop provides literally no relief, no individual makes any difference to the valuable outcome.[19] It would seem therefore that no individual has a reason to add his pint.

Nefksy argues that the problem can be solved by showing that the individual can be non-superfluously *causally involved* in collective impact without making a difference to that impact. She cashes out this idea of non-superfluous involvement by adducing examples of superfluous involvement. So, building on Parfit's case, suppose a group of individuals, A, recognising the needs of another group, B, each donate a pint of water and place it into a water tank so that it can be sent into the desert to relieve the latter's thirst. I then place a high-pressure hose connected to my own water supply into the already full tank, and blast water into it with such force that all of its present contents are expelled onto the ground and replaced with my water, before it is despatched into the desert. In such a situation, Nefksy points out, I am causally involved in relieving the thirst of group B, but my action is superfluous, and therefore not *helpful* (indeed it is wasteful). I therefore clearly have no reason to do it. In Parfit's standard case, however, in which the water tank is not already full, I am non-superfluous, in the sense that it is possible, before the tank is filled, that group A will fail to relieve the thirst of group B, and possible that it will fail due to an insufficient number of acts of the kind that I perform, even though the pint I add makes no difference, as the amount of water I donate to each individual is too small to provide relief.

[19] This is supposing for the sake of argument the outcome we are interested is providing the subjective experience of relief, such that, if the recipient cannot perceive a benefit, then no benefit has been received. Parfit wants to say the individual provides a real, though imperceptible benefit, Nefsky wants to say that the individual makes no difference at all to benefit. Whatever our view on the proper treatment of the Parfit case, it should be clear that it is possible to describe collective impact cases in which the individual's action makes literally no difference to the relevant impact, even if this is not such a case.

On this basis, Nefsky proposes the following specification for "helping":

> *Suppose your act of X-ing could be part of what causes outcome Y. In this case, your act of X-ing is non-superfluous and so could help to bring about Y if and only if, at the time at which you X,*
>
> *(*) It is possible that Y will fail to come about due, at least in part, to a lack of X- ing.*
>
> *Contained in this account are three conditions that are worth separating out. First, contained in the supposition that your act of X-ing could be part of what causes Y, we have:*
>
> *(1) It is possible that Y will occur. An act cannot be potentially part of the cause of Y if Y is impossible.*
>
> *Contained in (*) we have:*
>
> *(2) It is possible that Y will fail to occur, and*
>
> *(3) It is possible that Y will fail to occur at least partly as a result of there not having been enough acts of X-ing.* (Nefksy 2017 11)

Nefsky therefore gives the individual a reason to add his pint in Drops of Water – that his action will be *non-superfluously* causally involved in relieving thirst, in other words, that it will *help* to achieve that outcome. Nefsky's contention is that hitherto in moral discourse, we have standardly been working with an assumption that we cannot help to produce a morally significant outcome, without making a difference to that outcome. Her approach consists in challenging that assumption.

Even if we are willing to accept Nefsky's claim that we can help without making a difference, there is still some way to go before this distinction can form the basis of a solution to the problem. What remains to be shown is that it is helping, rather than difference making, that we should care about. In cases like Drops of Water, it feels, for most of us at least, like we are being presented with a falsidical paradox: we feel the result that I have no reason to add my pint just *must* be wrong, there *ought* to be a reason for me to add my pint. The problem is simply that we are unable immediately to see what it might be. Adducing the

reason that I thereby *help* to achieve the good outcome might therefore be considered a welcome solution to this worry. But arguably, there remains a more persistent problem: what to say to those who firmly assert that we have no reason to perform or refrain from performing the given action in a collective impact case, those who believe that individualistic rational choice theory gives us the *right* answer? Nefsky's account is arguably vulnerable to a renewed difference making challenge: why should I help (or refrain from helping) when helping won't make a difference?

We can easily imagine authors sceptical of individual negative obligations with respect to climate change mitigation offering this retort. The likes of Sinnott-Armstrong (2005) argue that because my action does no harm, I have no reason not to perform it. Nefsky replies that I have reason to refrain from helping to cause harm. The sceptic will say that this is moving the goalposts: he can simply deny that being non-superfluously causally involved in climate harms troubles him, given he makes no difference. Certainly, his actions are not *excessive* with respect to the outcome, it's not the case that they occur after the threshold for impact has already been met, as in the power hose case – but an act can be less than excessive without it being significant. Refraining from performing some carbon emitting act may prevent me from being non-superfluously causally involved in harm, but if I don't make a difference to that harm, how am I to attach negative weight to this consideration in my deliberation? To answer such a sceptic, it looks like Nefsky needs to provide an answer to the question of why helping matters.

Is Helping What We Care About?

Some of Nefsky's arguments are apparently intended to suggest not simply that difference making and helping are distinguishable, but that when it is possible to bring these two notions apart, it is helping rather than difference making that has, as it were, been our real concern all along. She does this by examining various non-collective cases where I seek to avoid some harm or to obtain some benefit, which seem to imply that difference making is not what we standardly value in everyday practical reasoning. In one case, say I wish to go to the supermarket. There are two routes that I might take, such that it makes no difference which I choose to follow. So, we might say, it *makes no difference* to the outcome of getting my shopping that I take route A, because if I didn't take route A, I'd just take route B instead. But, she observes, if I do take route A, I certainly have reason to do so. We can explain this, the thought goes, because taking route A will still be *non-superfluously causally efficacious* in producing the outcome of getting groceries. So, in this case, it isn't the fact that my act of taking route A makes a difference which I care about when deciding whether to perform it, but rather that taking route A *helps* to achieve my goal. Another case: I want my friend to receive a birthday card on his birthday tomorrow, so I need to ensure the card gets posted before the end of the day today. My housemate tells me: "if you don't have time to post the card in the morning, I'll post it this afternoon". Here, it looks like whether I post the card in the morning or not will make no difference to the desired outcome of ensuring the friend receives the card on time. But, argues Nefsky, it does not follow that I have no reason to post the card in the morning, since my action will still be non-superfluously causally efficacious in producing that outcome: before my housemate acts, it is still possible that the outcome will not be achieved, due to a deficit of acts of the kind that I perform (all above Nefsky 2012).

I would argue that neither of these cases gives us an answer to someone who is sceptical of the idea that I have reason to help when doing so will make no difference. The supermarket case seems to turn on a kind of equivocation. It may make no difference whether I go by route A or route B, but it certainly makes a difference *that* I go via route A or route B. In a collective harm case, my action makes no difference because whatever I do the outcome will happen anyway, *whether I want it to or not*. For example, let's return to the case in which I can either book a conventional taxi or a low-emission taxi, and I'm worried about whether causal involvement in climate change gives me a reason to choose one or the other. In this situation, I really have a third choice to consider: do nothing at all. The problem of collective harm arguably consists in the thought that even if I *do nothing at all*, climate change will be just as bad. In the supermarket case, meanwhile, if I do not go via route A, it is not as if I will then go via route B automatically. I still have to choose to do so. If I choose the implicit third option, and do nothing at all, it is very clear that this *will* make a difference: I won't get my shopping. So concern that one's action makes a difference is arguably still rationally operative in this case. Nefsky apparently conflates two distinct "differences" that might be "made": the difference that my action, whatever it is, might have on the outcome, and the difference in levels of effectiveness between different means of achieving an outcome.

In the birthday card case, we may question whether it is this idea of "helping" that really provides us with a reason to act. Imagine my rational psychology in this case. It does not seem likely I would recognise that merely helping to ensure the card arrives on time was my reason, rather, *making sure* the card arrived on time would most likely be my reason, in conjunction with a desire not to put my friend to any trouble. We can show this conjunctive reason is more realistic than merely the wish to help ensure the card arrives on time, by imagining a parallel case in which I care that the card arrives on time, but have no

particular concern that I should be the one to post it. For example, perhaps I work in an office with post collection. If I don't walk down to the post room in the morning, the office administration assistant will collect all the post for the office and walk it down to the post room in the afternoon. Picking up this card along with the rest will not put the assistant to additional trouble. In this case, if I post the card in the morning, I "help" to achieve the outcome of making sure the card arrives on time according to Nefsky's criteria. Nevertheless, it looks like I have no reason to do it; I'd be wasting my time.

Perhaps Nefsky wants to say that while helping does give me a reason for action in the birthday card case, it is a weak reason. She does note cases in which the particular circumstances make it the case that one's helping-based reason is weak. For instance, '[i]f Y is unlikely', she writes 'one can still satisfy the conditions for helping; one's reason to act might just not be as strong as it is in a case in which the chances of Y are closer to 50–50' (Nefsky 2017 21). But I would argue this response cannot plausibly be adapted to apply to the birthday card case. Here, we apparently lack a reason not because of the low probability of success, but because of the high probability that success is already assured. We could imagine that carrying the card myself even made things *worse* relative to the outcome: perhaps I know the assistant to be far more reliable than I am, while I might very well end up placing the card into the wrong post bag so that it isn't collected on time. Surely I cannot have a reason to act in a way that not only has a low absolute probability of success, but actually *decreases* the likelihood of my desired outcome (at least not a reason relative to that outcome). Nefsky cannot apparently treat this as a case in which her condition (2) is not met, however, because it is still *possible* that the assistant should fail to post the card, in exactly the same way that it is possible the housemate should fail to post the card in Nefsky's original case. The reliability of the housemate cannot have been in question in the

original case, as, if it were, it might be that my action made a difference, by increasing the probability of the outcome.

We might wish to respond on Nefsky's behalf, by claiming that she can account for the office case by making a minor alternation to her criteria for helping. Perhaps what we need to do is strengthen her second condition, so that where she has:

> *(2) It is possible that Y will fail to occur*

we would have:

> *(2*) The probability of Y's failing to occur is greater than zero, and greater than or equal to the probability that could be achieved as a result of my choosing to become causally involved*

This would mean that I do not help in the office case, and so would avoid the false-positive conclusion that I have a reason for action. And because the condition stipulates only that my action does not lower the probability of success, but not that it raises the probability, there could be cases in which the condition holds but in which I do not make a difference, which is what Nefsky needs to show – that these notions can be brought apart.

We may worry that introducing probabilistic notions into the specification at all, however, is going to bring unwanted complications along with it. Probabilistic notions typically come into play where access to information is limited (while there may be such things as objective physical probabilities, this is not the conception of probability we are typically applying when we discuss predictions of human behaviour). Thus when we say, in Drops of Water, that the probability of failure is greater than zero, we mean that from the perspective of an agent at some time prior to the tank being filled, given the information to which that agent has access, the outcome may not come to pass. If that agent had complete information about the psychological states of the other agents, and any local environmental conditions that might cause failures for reasons outside the agents' control, then the probability of the

outcome's occurring would be either 1 or 0. From a subjective perspective, a criterion that required one to estimate the probable impact of the outcome's chance of success might simple be impossible to apply. If I knew in advance that my action would increase the probability of the good outcome's occurring, then this alone would, it seems, give me a reason for action – it would mean that I *made a difference* to the outcome, if only in probabilistic terms. Yet if I make such a difference, then what we are faced with is not a "true" collective impact case, of the kind Nefsky is interested in. There could be cases in which the individual does not even make a probabilistic difference to the outcome, and yet there is a collective impact. Let us suppose therefore, that we are in such a situation: my action *cannot* increase the probability of the good outcome, as this would be to make a difference, and we are assuming that I make no difference. To know that my action was helpful, under condition (2*), therefore, I would need to be sure that choosing to become causally involved made, at worst, no difference to the probability of achieving the outcome. But given that*, ex hypothesi*, the information that the individual has access to about how future events will unfold is limited, it does not seem this is something that could be determined before the fact.

We could make the point more vivid in the following way: we might worry that, while my action cannot increase the probability of the outcome, my input *could* always make things worse. I could, for example, accidentally knock over the water tank as I add my pint in the Drops of Water case. Thus as my input cannot make a positive probabilistic difference without this ceasing to be a collective impact case, but I cannot rule out there being some probability, however negligible, that I make things worse, (2*) can never be satisfied. So I cannot help to bring about an outcome in situations in which I don't make a difference, on this view.

It might be remarked that the consideration that I could make thing worse does not apply to collective impact cases in which bad outcomes are avoided through individuals *refraining* from helping. If I am already performing an act, and several acts of this kind are together producing a bad outcome, how can it be that refraining from performing an act of that kind made things worse? In certain cases, this response is quite persuasive. In Harmless Torturers (Parfit 1984 80), for example, it is difficult to see how an individual's refraining from turning up the voltage dial could make things worse for the victim. I reserve judgement on whether condition (2*) could be satisfied in this case. But in a case like the taxi case (see Introduction), in which an individual has to make a decision about whether to avoid a negligible amount of carbon emission at some cost to himself, providing that he cannot make a positive difference to the probability of harm, it looks like we can always think of ways in which his decision could make things worse. The cost of choosing the lower-emitting option might be significant to me, and therefore it might have an impact on the way I live my life, which may, far down the line, prevent me from playing an important role in significant political action to combat climate change. The money saved might, with the effects of compound interest, be put towards some valuable adaptation project at some point in the far future. The probability of making things worse might be trivial, but it could still be greater than 0. Even if the suggested mechanisms through which individual action could lower the probability of a good outcome seem somewhat farfetched, the underlying point remains: a specification for 'helping' that required me to assess the probable impact my action would have on the outcome would not in practice be operationalisable, as this information is simply not available to me. Moreover, it now looks like, in order to have an expectation that my action will be helpful, and therefore to have a reason to perform it, I must *know* that it will make precisely zero difference to the outcome (in order to rule out

the possibility it will make things worse). It looks like a very odd result to claim that I have a reason for action in such a case.

Nefsky's proposed solution, therefore, seems merely to push the problem back a short distance. The fact that my action is not *wasteful* with respect to an outcome is not in of itself a reason to perform it. Moreover, as Nefsky herself recognises, even if we accept that the prospect of becoming a non-superfluous part of the cause of an outcome can provide us with a reason for action, we have still said nothing about the *strength* of the reason. It is part of the structure of many tricky collective impact cases, especially those that describe a real-world situation, such as climate change, that there are costs associated with refraining from becoming causally involved in harm. It is very difficult to imagine how we would weigh our reason to avoid acting in a way that is non-superfluous with respect to a harmful outcome against other practical considerations, such as our reason to avoid determinate costs. *Prima facie*, it might appear that the reason seems extremely weak, if we accept it is a reason at all. Many collective impact problems, however, surround situations in which we typically take ourselves to have very strong reasons for action. We standardly think, for example, that voting in elections is very important, even though our individual vote apparently makes no difference to the outcome of getting our preferred candidate elected. This lack of fit between the nature of the problem and the proposed solution might be taken to suggest that Nefsky's approach is misguided at a quite fundamental level.

The Joint Causation View

Nefsky, then, shows that while the consideration that by refraining from contributing to a collectively harmful outcome, I am helping to avoid that outcome, gives us some reason to refrain from contributing, that reason is very weak. Because in some cases that meet her definition of helping, we make things worse, it cannot be that helping alone, on this definition, gives one a reason to contribute. Her account could be amended to exclude the possibility that my choice not to become causally involved in the outcome makes things worse, but in doing so, the account reduces to an appeal to expected value. As Nefsky's response to Kagan (Nefsky 2011) showed, such accounts fail to solve collective harm problems as they cannot guarantee the expected value calculation comes out positive: perhaps the costs of individual action outweigh the value of increasing the probability of benefit, or the risk of harm.

Holly Lawford-Smith argues that individual contributions to climate change not only increase the probability of harm, they can actually be said to directly cause actual harm, meaning individuals can be regarded as wholly responsible for harm (Lawford-Smith 2016). She does so by positing the existence of "micro-thresholds" in GHG emission levels, at which emissions trigger certain harms. These are analogous to large-scale thresholds in the so called "damage function" – the pattern of expected harm that is predicted to occur as emissions, and correspondingly, global average temperature, increases. For example, changes in the global climate might disrupt the ecosystem of the Amazon rainforest, causing massive forest dieback, thereby eliminating a huge carbon sink and causing the pace of climate change, and the harms that go with it, to increase at a much faster rate (Lovejoy and Nobre 2018). Micro-thresholds are a similar idea on a smaller scale: the thought is that emissions on the scale of those produced by a smallish number of individuals might together trigger some discrete change in atmospheric conditions that produces some fairly

large discrete quantity of harm, by causing a particular weather even to play out in a more harmful way. In Lawford-Smith's hypothetical example, perhaps for every 1000 individuals who choose to fly from Australia to New Zealand, 10 additional people die.

Whether such micro-thresholds exist is an empirical question, which falls to science either to confirm or refute, but I do not here wish to deny that it may strike us as plausible. Even if they exist, though, one might think this gets us not much closer to difference-making. Assuming for the sake of simplicity there are only 1000 people in our universe under consideration. It is perhaps natural to think that only the 1000th person - the one who actually passes the threshold - causes harm, because she triggers the harmful event. Though we could, like Kagan (2011), appeal to the probability of being the trigger to give us a reason not to take the flight, we cannot say in advance that the expected harm associated with taking the flight outweighs the expected benefit of taking it, therefore that one ought not to take the flight.

Lawford-Smith resists this suggestion. Remaining for the moment in our 1000-person world:

> *The triggering of the micro-threshold and the subsequent death of ten people is counterfactually dependent upon all 1,000 individuals having chosen to fly. Each of those individuals is a difference-maker because without any of them having chosen to fly, the threshold wouldn't have been crossed.*
>
> (Lawford-Smith 2016 73)

Thus on this account, all 1000 people cause the harmful event – a case of *joint causation* – and should be regarded as *jointly* responsible for that action. But of course, it is unrealistic to restrict our consideration to just 1000 individual agents in relation to a particular micro-threshold. What should we say of cases that are not so restricted? The scenario then becomes one of causal over-determination. When considering the 1001st flight, she claims:

> *[E]ither all 1,001 routes jointly caused the ten deaths (joint causation), or 1,000 flights caused the ten deaths and pre-empted the 1,001st flight from causing anything (pre-emptive causation). But on the latter, now factor in that there are many flights being taken each day and imagine that the threshold is always 1,000; then, even being the 1,001st relative to that one threshold doesn't mean your action doesn't make a difference, because it might become the first of 1,000 flight-takings jointly necessary to triggering the next threshold.*
>
> (Lawford-Smith 2016 74-75)

Thus, for every emissions-causing act, one is always a member of a group that jointly triggers some threshold. She argues this partly by analogy with a solution to another paradox of difference-making, the voter paradox. The voter paradox consists, on the one hand, in the idea that, whether one votes or not one will have no effect on the outcome, because there is only one possible outcome on which one's vote makes a difference – the outcome on which one's vote is decisive, breaking an exact tie – and that outcome is vanishingly improbable given an election with a large number of participants. Yet, on the other hand, if many people reason thus, their abstentions are in combination very likely to have an effect on the outcome. The proposed solution (credited to Richard Tuck) lies with the thought that one's status as tie-breaker is only salient if votes are cast diachronically and one is chronologically the last to vote. As Tuck notes, elections in the Roman Republic were in some sense conducted in this way: each Roman tribe took it in turns to vote and when it became clear that a candidate for tribune had achieved a majority, voting stopped whether or not all of the tribes had cast their votes (Tuck 2008).

If votes are cast synchronically, meanwhile, there is no distinguishing causally efficacious votes from "pre-empted" votes, and we therefore must say the candidate was elected by all those who voted for her *jointly*. In other words, we abandon counterfactual dependence as a criterion for causation, to allow for 'redundant causation'. We may say, following Tuck, that if for example 10,000 people voted for a candidate, and 4000 votes was the threshold

necessary to elect the candidate, each voter has a 2 in 5 chance of having been causally efficacious. Given this chance is fairly significant, individuals have good reason to vote. Similarly, in the case of climate change, if we accept the existence of "micro-thresholds", each of those who chose to take flights have a high probability of having been causally efficacious relative to the given threshold, and, the thought goes, near-certainty of having been causally efficacious relative to some threshold or other.

The analogy with the voter paradox only seems to be helpful so far, however. In the voting scenario, if we cast our votes diachronically, then it is literally the case that the n+2nd voter does not make a difference (where "n" is the threshold number of votes). It is only by stipulating that votes are in fact cast synchronically that we get joint difference-making. In the climate change case, meanwhile, it appears Lawford-Smith wants to say that there is *in fact* some individual whose emissions surpass the given threshold, but the causal connection between that individual and the threshold effect is 'epistemically opaque' to us. If this is true, it looks like she has to regard our contribution to emissions as taking place diachronically. But if contributions are genuinely diachronic, then she cannot appeal to the solution to the paradox offered by Tuck. If emissions-causing actions are regarded as occurring in sequence, then there is some individual that triggers the threshold, and if there is some individual that triggers the threshold, then that individual should be regarded as the cause of the harm.

In an election, it is, as Tuck writes, 'perfectly reasonable to say that my vote may bring about the result even if it is not pivotal' (Tuck 2008 36), because precisely those voters that elected the candidate have to be considered the cause of the candidate's election (what else could be?) even though counterfactual dependence may not hold if there are a large number of votes in excess of the threshold. Micro-thresholds in GHG emissions would not be like this; it is an artificial imposition on nature to view them as operating in such a static

way. In the election case, there are a fixed number of votes which go towards a one-off causal effect, and so it makes sense to include all of them as the cause of the candidate's election. Yet in the case of the 1000 flights threshold, even if all 1000 people had refrained from taking the flight, the threshold could still have been surpassed. Another 1000 somewhere else could have done the job, and if not them, another 1000. Thus, the size of the group of potential joint-causers is not naturally bounded. If my reason to avoid emissions is my chance of triggering a threshold, and this is calculated as the size of the threshold (1000 flights) divided by the number of people that actually bought flights, this chance is arbitrarily small as the latter group is arbitrarily large.

Perhaps the clue as to which acts count as contributing to the threshold is the very fact they happen at the same time. But this is a difficult idea to apply. In the case of voting, we say that votes are cast synchronically because it is a formal property of the counting system that chronological order does not matter. A number of different counters could finish their piles at different times and it only when the results of all the piles are added up that we get the overall results, so no individual can be viewed as breaking a tie at a particular moment in time. We could even imagine the voting process being done through pressing a button on a screen exactly simultaneously. When it comes to GHG thresholds, however, it appears that if Lawford-Smith wants to say that we cannot - simply as a consequence of available measuring technology - know when a threshold is surpassed, it must also be the case that it is only for epistemic reasons that we cannot strictly order GHG emission events to determine which event triggered the threshold. If we grant that only one person is the trigger, this would reduce the probability of being a trigger in any particular emission activity. If we follow Kagan (2011), we might consider the probability of being a trigger to be sufficient to ground a duty not to emit. Nefsky, though, gives us a strong case as to why Kagan has insufficient grounds for his conclusion: 'there is no guarantee that expected

utility will come out negative in every triggering case (Nefsky 2012 369). If, on the other hand, Lawford-Smith denies that any particular people break the threshold, then it is difficult how she can maintain that anyone makes a difference.

Joint Difference-Making and Responsibility

When it comes to the question of how responsibility for triggering a threshold is to be distributed among the group of joint-causers, the notion of joint causation becomes even more problematic. Tuck claims that in voting 'each vote carries the full causal responsibility for bringing about the result' (Tuck 2008 41). This thought is essential to his argument that rational choice theory can indeed recommend voting. The idea is that 'one can assign a measure of utility to an action equivalent to the consequences which it brings about' (Ibid.). Thus the value of voting is just the value of whatever good one hopes will be achieved by the election of one's chosen candidate. He notes that one might at this point worry that this leads to an absurd result – that if each of 1000 voters assesses the utility of the outcome at 100 units, the expected value of the outcome would be 100,000 units. This result can be avoided, he claims, by observing that it involves double counting: each individual assesses the global expected utility of the outcome at 100, a result that would itself be arrived at by summing the expected utilities for each of the beneficiaries.

Whatever we think of this view we may doubt that the same intuitions apply when we are considering not expected benefit, but harm. If a group of individuals 'jointly' causes the triggering of some catastrophic threshold-effect, such as the melting of the Siberian permafrost causing the release of millions of tons of methane that was trapped inside it, it does not look correct to say that each individual is causally responsible for all of the harm associated with this macro-level effect. Tuck and Lawford-Smith would perhaps argue that the apparent absurdity stems from an illegitimate conflation of causal and moral

responsibility. Yet Lawford-Smith in particular seems to acknowledge that her argument might indeed have this odd result:

> *Causation and responsibility are not necessarily proportionate, so being blameworthy along with 999 others for the death of ten people does not mean being a thousandth responsible for ten deaths (or a hundredth responsible for one death). Each individual might be fully responsible for all ten deaths, given that her choosing not to fly would have been sufficient to those deaths being avoided.*
>
> (Lawford-Smith 2016, footnote)

In this footnote, I take it that Lawford-Smith is limiting her attention to the 1000-person world. She here links blameworthiness, which is to say backward-looking moral responsibility, with counterfactual dependence. Leaving the 1000-person world would then mean counterfactual dependence would no longer hold. What then should we say about moral responsibility? Lawford-Smith wants to say that even if we do not end up as part of the 1000-person group of joint-causers, we would then roll over to the next threshold and become a joint-causer in that group. Does this mean that, in effect, although the individual cannot be considered morally responsible for any "10 deaths" (any threshold-effect harm) in particular, she is responsible for some "10 deaths" or other? This intuitively looks excessive, and it is difficult to rationalise. Given that it is not true that had the individual not contributed, the harmful event would not have happed, it is not clear why we should regard the individual as morally "on the hook" for the harm caused by the triggered event.

Moreover, we might worry that the view radically over-generates. Lawford-Smith apparently wants to draw a strong connection between causal responsibility and moral responsibility. Now, it is not actually true that individuals cause GHG emissions in neat 1000-person cohorts. Presumably, every GHG-emitting act going on at the present moment is causally connected to the passing of innumerable micro-thresholds at some point in the future. One individual's emissions are part of the cause of the state of future climatic conditions at a

given moment, and those conditions are part of the cause of future climatic conditions at succeeding moments. On the joint-causation account, there is no apparent way of individuating these causal connections, so that particular sets of people are connected to some thresholds but not others. Therefore, if my emissions are part of the cause of one threshold being surpassed, they are also part of the cause of the next being surpassed, and the next. Thus it seems difficult to avoid the conclusion that on the joint-causation account, every single individual act that causes emissions - like the choice in the taxi case - is causally responsible for the whole of future climate change. From a metaphysical standpoint, this is not necessarily a *reductio* of the view, but it does stretch the notion of causation rather thin. More importantly, though, if Lawford-Smith does indeed intend to relate causal responsibility to moral responsibility, we might fear the view stretches the concept of moral responsibility to breaking point.

Lawford-Smith only claims that her arguments support the conclusion that we are very likely to make a morally significant difference when we contribute to GHG emissions. They are not intended to support the claim that there is a directly proportional relationship between the amount of GHGs we emit and the amount of harm we do. Thus the account cannot give us the idea of *liability* for contributions to climate change: it cannot give us differentiated responsibility for the amount of harm that each individual causes, nor can it distinguish between liable and non-liable contributions on the basis of severity. At this point, we might wonder what exactly it is that the account gives us. If the connection between causal and moral responsibility is so vexed and obscure, what does the account imply that we should do in a case like our initial taxi case? Lawford-Smith claims that her argument nevertheless has practical implications: for example, she suggests we should regard our individual difference-making as producing only 'positive obligations' (Lawford-Smith 2016 78), meaning contributions to GHG emissions do not generate what Honorés

calls outcome responsibility, or backward-looking responsibility, only forward-looking responsibility. But there are ways of grounding such obligations without the assertion that individuals are causally and morally responsible for triggering significant harm, and the joint-causation view does not appear to be the most obvious or direct way of doing so.

A final suggestion Lawford-Smith makes is that we can incorporate a distinction between 'subsistence emissions' and 'luxury emissions' into her account, in order to make it more directly action-guiding in cases like the taxi case. The suggestion is that by means of this distinction, we 'would be able to say that only certain kinds of emitting acts are even *prima facie* wrong' (ibid.), which, if true, would allow her to overcome the suspicion that her account over-generates, implying the existence of wrongs that we would not wish to regard as such – implying, in other words, that all contributions to GHG emissions are equally wrong. She could then claim to be able to pick out negative obligations that apply in some cases and not others. Yet this suggestion is perplexing. On Lawford-Smith's framework, if a case of luxury emissions is wrong, it is wrong because it is harmful. But a case of subsistence emissions may be harmful in just the same way. It seems most natural here to say that both cases of emissions are *prima facie* wrong, but that the wrong is mitigated by the fact the act was necessary in the case of subsistence emissions, and is thus not wrong all-things-considered, while it remains wrong in the case of luxury emissions. Yet if this is admitted, our inability to know the *degree* of harm an individual's emissions cause becomes problematic. We would not say the fact some action was necessary for my subsistence justifies all and any wrongdoing that might be associated with it. If, in order to avoid starvation, it was necessary for me to kill 10 people, we would still wish to say that I do wrong by killing 10 people, even though we might also grant that my behaviour was excusable, although not justified. This would be a case of tragic choice. Thus, unless we have a grasp of *how much* harm individuals cause, we cannot say with certainty that

wrongdoing is mitigated by the fact it stems from the need to provide for one's own subsistence

This point is brought out more forcefully when one considers that, in his original article (Shue 1993) on subsistence emissions and luxury emissions, Henry Shue did not apply this distinction in deciding which emissions were acceptable on an individual level, but only to questions of international burden-sharing. The distinction is apparently less suited to application at the level of first-personal practical reasoning. It seems to have been conceived, in essence, as a rough-and-ready divide between emissions related to broad economic sectors - to food production on the one hand, and luxury consumer goods such as expensive cars on the other. This coarse-gained approach is relatively unproblematic when considering whole economies. From the perspective of the individual, however, it is not so easy to draw the same kind of line. Our food purchases can be made with a very great variety of levels of concern about related emissions. Should we buy only foods that are produced in our own country? Or in our own local area? Should we boycott foods produced through intensive farming techniques? All these choices are going to come with costs attached to them. Those who are materially better off will be able to afford to comply with more stringent principles; it may be reasonable to have lower expectations of the less affluent. Lawford-Smith's account does not seem to be able to assist with this kind of fine-grained decision-making.

Similarly, our concept of need in the case of individuals is generally considered to be more complex than simply "that without which the individual will die". As Adam Smith noted, need has a social component: an Englishman in Smith's time needed leather shoes, because it was in Smith's view impossible to be seen in public without them except with great shame, while in contemporary France, wooden clogs were considered sufficient (Smith 1776 818). Joseph Raz has suggested the idea of 'personal needs', arguing that once

we are 'set upon' (Raz 1988 157) certain paths in life, this can generate a claim of need for those things necessary to continue on that path. In his example, pianists may 'lose the life they have' if their fingers are broken. For certain individuals, the same might go for losing access to goods that others might regard as luxuries – access to pianos, for instance. David Wiggins has suggested we think of needs as states of affairs whose obtaining is necessary for some 'unforsakable end' to come to pass, where an unforsakable end is an end without whose fulfilment 'a subject will be seriously harmed, or else...will live a life that is vitally impaired' (Wiggins 2005 31). Harm is itself to be understood in relation to 'a minimal level of flourishing' (Wiggins 1998 13). Thus again, on this account needs go beyond mere subsistence and, moreover, are necessarily judged in comparison to one's social circumstances. A simple distinction between subsistence and luxury emissions is not going to help us to make judgments about which emission-causing activities form part of an individual's *needs*, as this idea plausibly implies some standard of what is socially *possible*, which in turn depends upon the costs it is reasonable to impose on others. Assessing the *level* of harm, or of risk, that our putative needs impose would ideally form part of the process of establishing this standard. Yet this is precisely what we are told is impossible: on Lawford-Smith's account we can know that we make a difference to harms, but we cannot know what difference we make.

Thus it looks like Lawford-Smith's view is either deeply counter-intuitive or practically inert. Her view of causal responsibility seems to give us the wrong conclusions if we attempt to use it as the basis of an account of moral responsibility. Yet if we do not do so, and view our "difference making" as generating only "positive obligations", then it is hard to see what the theory has added. Even Sinnott-Armstrong, whose 2005 paper caused serious controversy with its extreme antipathy to the view that individuals had obligations in relation to mitigation, acknowledged that individuals had 'positive obligations', such as the

obligation to lobby their governments to impose coercive measures to promote collectively rational behaviour, the traditional solution to collective action problems. Although Lawford-Smith's focus was the question of whether the individual 'makes a difference', this is not the most important point for our purposes. Even if we grant, for the sake of argument, that Lawford-Smith is right that the individual makes a difference, the problem remains that her explanation of *why* the individual makes a difference does not inform our moral reasoning in cases like the taxi case, and it does not help us with the question of how outcome responsibility is to be assigned.

Individual Direct Causation: Linear and Chaotic

Broome's account seems, at least initially, to be much more directly action guiding than either Lawford-Smith's view or Nefsky's view: he claims that individual contributions to GHG emissions are unjust, implying that they are, as far as possible, to be avoided in all circumstances. Recall from Chapter 2 that Broome offers seven conditions that are in combination supposed to show that individual contributions to GHG emissions constitute injustice. They are: i) the harm is the result of an act (rather than an omission), ii) it is serious, iii) it is non-accidental, iv) it is not compensated, v) the emitter benefits from her harmful activity, vi) the harm is not reciprocated, and vii) the harm could easily be avoided (Broome 2012 55-59). In *Climate Matters*, Broome was not always clear in distinguishing two separable claims, that individuals actually do harm through their emissions, and that there is an expectation of harm attached to individual emissions. Only the latter of these claims was supported by his argument, and it is not obvious that one necessarily has a duty to avoid actions that carry some expectation of harm. Thus, for example, if I choose to go for a bike ride, pedestrians on my route are placed at greater risk than they would be if I choose not to go for a bike ride. We would not for that reason alone, however, say that I had a duty not to go for a bike ride at all, so long as I took reasonable precautions.

Perhaps acknowledging concerns of this kind, Broome has in more recent work given a more precise account of how he takes the causal relationship between individual contributions to emissions and climate-related harms to operate. On the one hand, he takes it to be at least possible that there is a linear causal pathway between emissions and harm, such that marginal emissions always cause marginal actual harm at any scale of emissions. If this is right, then, Broome thinks, we can say determinately that individuals cause harm through their emissions, and not just that their emissions carry an expectation of harm. The

more cautious statement of this view in (Broome 2019) marks a departure from earlier work, where he stated very directly that 'the harm done by greenhouse gas emissions is proportional to the quantity of emissions' (Broome 2016 160), but qualified this remark by saying that he was in reality talking about expected harm.

His current view is apparently that a portion of the actual harm caused by one's emissions may increase in direct proportion to the quantity of emissions produced, and for the rest, what actual harm any particular emissions-causing event does will vary wildly, some causing vast amounts of harm, some very little, and others actually producing benefits. Thus, on this second, chaotic causal pathway, emissions trigger harms in unpredictable ways, and harm is not necessarily proportional to the quantity of emissions. This latter description of the causal pathway appeals to the so-called "butterfly effect": the idea that small changes to a complex system can produce very large-scale consequences. The further claim is then that because we cannot predict the consequences of any particular emissions-causing event, but because we know that, at a large scale, more GHG in the atmosphere leads to more harm, we should treat the expectation of harm of any particular emissions causing act as sufficient reason to avoid it. Specifically, although the social cost of carbon is calculated at the scale of billions of tonnes, it is still reasonable to treat the proportional social cost of just a few kilograms of carbon as an estimate for the amount of harm emissions of that magnitude will do. We will examine in turn the implications of each of these characterisations of the causal pathway between emissions and harm .

Linear View

The claim here is that some of the harms of climate change are not triggered at particular thresholds, but rather are continuous, so that there is a directly proportional relationship between marginal GHG emissions and marginal harm "all the way down", as it were.

Broome points, for example, to 'the steadily increasing difficulty of getting water in some parts of the world' (Broome 2019 118). Broome refers to the fact that rising temperatures will cause water tables to drop, making it progressively more difficult for people to access water for drinking and irrigation. Elsewhere, he refers to the fact that 'the gradual rise in sea levels caused by climate change will steadily erode the land' (Broome 2016 161). Molecules of GHGs absorb energy and therefore, as Broome points out, every molecule of GHG produces a heating effect (for as long as it is in the atmosphere). Water tables can be expected to drop in proportion to rising average temperature. Therefore, any emission-causing act has some negative impact, however slight, on falling water tables, and therefore on harm, or so the argument goes. Falling water tables are only intended to be one example, but I shall continue to use the example for convenience's sake.

There are at least two potential problems with the claim that individuals cause harm in this linear way. For one, it is plausibly false that the individual's emissions cause a drop in the water table, let alone any attendant harm. It does not obviously follow from the fact that all emissions produce a heating effect that, if the water table drops in a given territory, any emissions event anywhere in the world can be judged to be the cause of a tiny part of that drop. The average level of a water table in a given region over a given period is determined, in part, by the average temperature in that region over that period. But one might think it is a kind of category mistake to think that because molecules of carbon produce a local heating effect, they thereby cause the regional average temperature to have a certain value over a given period, and thereby cause the level of the water table to have a certain value at a particular time. The fact that some molecules of carbon dioxide cause a heating effect in a certain location is rather partially *constitutive* of the fact global average temperature has a certain value. This is different from saying it causes it to have that value. Thus, it is arguable that it only makes sense to say that a group of emitters together cause the water table to

drop, but not that any individual causes it to drop. I accept, however, that this form of reasoning involves potentially controversial empirical assumptions which it may be possible to resist.

For another, it is plausible that minute changes in the water table are too small to constitute harms. This second point presents a stronger challenge to Broome. It is easy to see why one might think a drop in a water table of some tiny fraction of a centimetre could not constitute harm. Say we take the form of harm under consideration to be the additional effort it would be necessary to exert when digging a well before the aquifer was reached. Even if it is true that, at the time of digging, the water table is a tiny fraction of a centimetre lower than it would have been had a single individual not emitted, it is not plausible that this constitutes any more effort: plausibly, when digging a well, there is always going to be some margin of overshoot when deciding how deep to dig, and it is vanishingly improbable that the drop in the water table would take the level just beyond that margin of error.

Broome would likely respond that this form of reasoning simply reopens the debate about whether there are imperceptible harms, to which he believes Derek Parfit has already provided a basically adequate answer. Broome thinks we can bypass quasi-metaphysical arguments about whether the concept of imperceptible harm makes sense by treating the thesis that there are such harms as the conclusion of a reductio argument. For this argument, the sorites paradox is used to generate a contradiction on the assumption there are no imperceptible harms: the contradiction that digging to a depth that would plainly involve more effort would involve no more effort. One problem with this approach might be that it treats the paradoxical nature of the paradox rather flippantly (as, for example, Nefsky 2019 points out). The thesis that there are imperceptible harms can also be used to generate counterintuitive conclusions; this is precisely why it is a paradox. We can indeed run parallel reductio arguments which purport to falsify the existence of imperceptible

harms – we could generate the contradiction on the fact that the subject's situation between two increments is in no respect worse (having had to expend no more effort), and yet she is supposedly harmed.

We can do without this argument, however. Note that the claim the well-digger is in no respect worse off does not depend upon her reports, or her ability to discriminate between states. It simply depends on the idea that when digging a well, it is impossible to dig at micro-level precision, and so the chances are that the subject will expend very slightly more effort than she strictly needs to, whether or not the water level is a fraction of a centimetre lower. It might be countered that a fractionally lower water table makes the subject worse off in that it changes her options for the worse: hypothetically, when the water level is higher, the amount of effort required to dig to that level is lower, therefore, the subject has the option of expending less effort (even if in practice she does not use it) and she loses that option when water level is lower. But I would simply deny that she has any such option. Because, as already stated, it is impossible to dig to micro-level precision, the subject cannot avoid overshoot. An option one lacks the capacity to avail oneself of is no option at all. Thus it is just implausible that digging the well to the depth required if x people emit at some earlier time will involve any more effort than digging to the level it would be if x+1 people emit at some earlier time.

To be clear, the important question is not whether the water table falls for any incremental temperature change. It is indeed incoherent to suppose that it does not, since as Broome points out, if at temperature increment 2 the water were no lower than at increment 1, and at 3 no lower than 2, and so on to increment 1000 or 1,000,000, then the water table would never drop, a hypothesis falsified by the fact the water table in reality does drop. Our question is rather whether tiny changes in the water table constitute harm. It is plausible that, as Richard Yetter Chappell puts it, there is no such thing as 'ontic vagueness', i.e. 'the

world is precise and determinate in all fundamental respects' (Yetter Chappell n.d. 4). Vagueness would seem to be a semantic phenomenon, a manifestation of the fact that the extensions of certain predicates are not precisely delineated in language, not that the physical phenomena they designate are somehow themselves fuzzy at the boundary. Baldness is not an 'objective property' (ibid.); its extension is vague because it is a loosely defined semantic convention. Heat and the level of a water table, meanwhile, are both physical quantities, thus it is indeed implausible to suppose it could be indeterminate whether a change in heat or water level has occurred, given a sufficiently complete description of the circumstances.

It is not, however, implausible that our interests are vague, or that one's level of wellbeing could be vague. Yetter Chappell is wrong to claim that 'fundamental goods are not themselves vague' (ibid.), or at least the claim is under-argued. He makes reference to 'whatever natural features are of fundamental moral significance' (ibid.), and this itself is telling: it indicates he assumes fundamental ethical significance must attach to natural properties. It may be, however, that fundamental ethical significance attaches not to natural properties (or not only to natural properties) but (also) to social or conventional properties, which may be vague. Ethical properties might attach directly to conventional properties, and not to whatever natural properties underlie them.[20] Ethical properties would then be "ontically vague", in that there would be no underlying dimension which could be used to sharpen them.[21]

Why should we think ethical properties are like this? Many theories of wellbeing take wellbeing to have a subjective component. For example, some people think the satisfaction

[20] David Lewis uses the term 'natural property' to mean properties 'whose sharing makes for resemblance, and the ones relevant to causal powers' (Lewis 1983 347).
[21] This is a loose usage of the term "ontic vagueness", as some may wish to reserve the term "ontic vagueness'" for vagueness "in the world", and deny that ethical properties could be said to exist "in the world" if they attached directly to conventional properties in this way.

of desires is an important part of wellbeing. The satisfaction of a desire has an objective and a subjective aspect: there is the desire, which is expressed in concepts, and there are the external facts which constitute its satisfaction. Whether my desire for a loving relationship is satisfied is a matter of whether love is the appropriate description for the state of affairs in which I find myself – a question which is in part conceptual. Even philosophers who hold that wellbeing is constituted by an objective list of goods (see eg. Parfit 1984 493) describe those goods by means of vague predicates. Whether someone is in possession of some good can therefore also be viewed as having a subjective component - whether "having such-and-such a good" is the correct description of what is going on – which again is a question of how certain concepts ought to be deployed.[22] If we think that loving relationships are an element of wellbeing, then the fact the concept "loving relationship" admits of borderline cases means that wellbeing could indeed be considered an instance of "ontic vagueness".

Views of the elements of wellbeing are positions in normative ethics, which may be underpinned by different positions in meta-ethics. Thomas Hurka's view that significant achievement is a key element of wellbeing, for example, is underpinned meta-ethically by a kind of Aristotelian naturalism, whereby whether an achievement is significant is determined by whether is contributes to one's perfection as a human being (Hurka 1996, 2010). If we have this kind of view, then Yetter Chappell's reference to 'whatever natural features are of fundamental moral significance' is not empty, and his claim that these features cannot in themselves be vague is plausible. Though we may identify those properties that contribute to the perfection of human nature with vague predicates, we can view these predicates as a kind of shorthand for whatever natural property, close to the

[22] Elson (2017 346) notes another example: if my good is constituted by what I would choose under ideal circumstances, and if those ideal circumstances are indeterminate in character, then this indeterminacy might create corresponding indeterminacy in what constitutes my good.

extension of the natural language predicate, is in fact partially constitutive of the perfection of human nature. But we need not have this kind of view. We may be value non-naturalists.

What does the possibility of evaluative vagueness mean for putative small harms in the case of climate change? Take the question of whether a micro-scale strip of land lost to the sea constitutes harm. It is plausible that whatever fundamental values are served by ownership or use rights over a certain piece of land, none of them is frustrated when only an infinitesimal fraction of that land is lost to the sea, because small losses in land do not correspond to small loses in underlying value. Say we take a piece of land to be valuable to a given individual because of some special relationship with it (it being her homeland, etc.). The loss of a microscopic strip of it does not make it lose any of that quality, or offend against the value of that relationship. Thus, though it is incoherent to suppose that no individual contributions to GHG emissions correspond to real temperature change, and or that no such individually-caused temperature change corresponds to an additional volume of liquid being added to the ocean, it is nevertheless highly plausible that no individual causes a rise in sea level that makes anyone worse off, or damages anyone's interests. The same is plausibly true for other continuous effects of climate change. Thus, Broome's claim that there may be continuous climate harms "all the way down" for any quantity of marginal emissions is insufficiently supported.

The Butterfly Effect and Expected Harm

Broome may not be greatly troubled by the conclusion that emissions do not determinately do actual harm at a small scale, however, as he presents this as an incidental and admittedly speculative component of his view. At the same time, he thinks that most, if not all, of the expected harm associated with an individual's emissions follows a causal pathway characterised by the 'butterfly effect'. This description clarifies Broome's early account in

an important respect. Broome is not arguing, as it originally seemed from his work, and as was picked up by, for example, Elizabeth Cripps, that the individual, through her emissions, does a small amount of harm to a large number of people. He therefore bypasses all the questions of whether small effects really count as harms at all, and whether they aggregate up to serious harm. Rather, Broome thinks that there is a significant probability that any given emissions-causing action will result in actual harm, and there is a significant probability that the harm will be serious. There is also, however, a significant probability that it will do good, or that it will do some combination of harm and good. It will always, however, have a negative expectation of harm, because on average, GHGs are harmful with respect to their forcing effect on atmospheric temperature.

As Broome acknowledges, then, '[s]ince imposing expected harm is imposing a risk of harm, the question is whether justice prohibits you from imposing a risk of harm' (Broome 2016 161). Why should we think this is so? Broome 'leans towards' the view that 'as well as the duty of justice not to harm people, there is another duty not to impose a risk of harm on them' (Ibid.). Not even all harms are injustices, however, so *a fortiori* not all acts which carry a risk of harm are injustices. As suggested earlier, even obviously innocent activities like going for a bike ride while taking all the normal precautions impose some additional risk of harm on the bystanders one passes. The question is then is what additional conditions have to be met to render the imposition of risk unjust? Broome writes, '[j]ustice requires you not to harm other people, at least not for your own benefit' (Broome 2016 161), and implies that the same principle applies to expected harm. But this principle gets us no further, as the case of carefully riding a bicycle down the street, which looks like an obviously unobjectionable act, is precisely a case of imposing risk for one's own benefit (assuming one is riding the bicycle for fun).

In more recent work, responding to Sinnott-Armstrong's case of the 'joyguzzling' driver (that is, driving a gas-guzzling car just for fun), Broom says the following:

> *Whether or not you ought to joyguzzle may depend on various things. No doubt joyguzzling brings some benefit to you, and this benefit may be worth more than the $1 of expected harm it brings other people. On the other hand, it may be that you ought not to expose people to even a small expectation of harm just for your own enjoyment.* (Broome 2019 115)

Where in *Climate Matters,* Broome claimed that one had a strict duty of justice to refrain from contribution to GHG emissions, Broome now states that whether one ought to joyguzzle depends on whether joyguzzling has a positive expected value. Whether this is so would depend on the size of the benefit attached to the carbon-intensive activity under consideration, and the social cost of GHGs. If the estimate of the social cost of carbon Broome adopts for the sake of argument – $40 – is accurate, this means the joyguzzler only has to justify a dollar of benefit, which she easily does. Even higher estimates for the social cost of carbon would leave joyguzzling in an appealing price bracket. Broome's target is now "denialism" – his provocative name for the claim that individual emissions make no difference. Individuals do make a difference, he argues, because they make a difference to expected harm. This gives them a reason to refrain from contributing to emissions, a reason of goodness. As he acknowledges, however, there are many more effective ways to do good, so our reasons of goodness to refrain from emissions are very weak. When we consider the question of responsibility, it does not seem Broome, in the final analysis of his most recent work, gives us a reason to hold individuals responsible for contributions to GHG gas emissions, as they do no wrong by their emissions, at least in most cases. The collective harm problem therefore remains unresolved.

Offsetting

Offsetting as Collective Harm

Despite this apparent softening of his position, perhaps Broome may still wish to maintain that it is wrong to impose a risk of harm if the cost of avoiding such harm is low. As we rely on fossil fuels for many important goods in our daily lives, this claim would be much more difficult to justify if reducing our emissions to net zero required us to stop using fossil fuels altogether. The appeal to low cost therefore relies heavily on the further claim that it is cheap, easy and effective to purchase carbon offset. This claim is problematic. The effectiveness of various offset schemes has been questioned empirically, so much so that Kevin Anderson has stated '[o]ffsetting is worse than doing nothing' (Anderson 2013 7). Concerns include the idea that offsetting projects which support development may actually lead to an increase, rather than a decrease in emissions when their effects are considered over a sufficiently long timescale, as increased overall economic activity will likely lead to increased fossil fuel demand. Furthermore, reliance on offsetting as a means of tackling climate impacts locks in dependence on fossil fuels. In other words, by buying carbon-intensive services today – flights, for example – one contributes to the continued expansion of the aviation industry, and thereby the greater availability of flights in the future, again potentially leading to higher overall emissions when considered over a long enough timescale.

To the first of these points, Broome might respond that the problem here is not intrinsic to offsetting; rather, he might say, a few bad apples in the offset market are bringing the entire concept into disrepute. To the second, he may respond that we simply need to make sure we purchase additional offset to compensate for these side-effects, as they have likely not been factored into the offset price. First, it's important to note the interdependence of

these responses: if offsetting can really do more harm than good, then by deliberately overestimating the amount of offset one would be required to purchase to undo the negative impacts of one's emissions, one would just be digging oneself a deeper hole. On the first point, we can concede that Anderson's target in particular is the Clean Development Mechanism (CDM) under the UNFCCC. Projects accredited under this initiative are specifically focused on development; we might therefore arguably have less cause for the kind of concerns Anderson raises with respect to voluntary schemes marketed primarily towards individuals, rather than state parties to the UNFCCC. Though one suspects that many offset providers would balk at the suggestion that the most effective forms of offsetting are those that contribute *least* to development, given that most tout the promotion of development as a co-benefit, it is perhaps possible to find schemes that have minimal side-effects in terms of contributing to carbon-intensive economic growth. A number of schemes, for example, involve the provision of fuel-efficient rocket stoves to households in developing countries that previously cooked on trivets over open fires. This represents such a minor change to household management that it is difficult to imagine it would engender serious changes to these households' prosperity and concomitant fossil fuel consumption. Although there are concerns about the additionality of such schemes, Broome may be correct that it is not impossible at least some of them have done their sums correctly.

The second point, however, is not so easily dismissed. It gets to the heart of the problem with views, like Broome's, that take the individual's primary duty to be avoiding doing harm through their personal emissions: individual offsetting itself reproduces collective impact problems. For one thing, because the impacts of the initial emission-causing behaviours are at some level essentially collective, they cannot in principle be mitigated by individual offsetting. My purchasing a seat on a flight arguably makes no difference to the scheduling

decisions of airlines. If we had to quantify the degree to which, by purchasing a flight, I cause future emissions by contributing to the expansion of the aviation industry, it would be difficult to find grounds to assign any figure other than zero. But, we know, the more people who fly, the more economic incentive there is for airlines to lay on additional routes, and to lobby governments to authorise the opening of additional runways. This collective effect does not arise directly from my decision to fly, but from that decision in combination with the market behaviour of millions of other consumers. The idea that I could simply factor this impact in to the amount I am obliged to offset is fundamentally misconceived. Such impacts are therefore systematically neglected by Broome's individualist approach.

Offsetting and the Butterfly Effect

Furthermore, there is arguably injustice embodied in the idea of cancelling one's emissions by, effectively, paying others not to emit. Elizabeth Cripps likens this to the idea that one could compensate for the harm of disturbing one's neighbours with a loud party, by paying the neighbours on the other side not to throw another party they were planning, thus reducing the net volume of noise to the same level as it would have been had one's own party not taken place (Cripps 2016). Broome denies this equivalence, arguing that because the harm done 'is determined through one quantity, the global concentration of greenhouse gas' (Broome 2016 159), offsetting is not like doing harm in one place and then attempting to compensate for it by preventing some other harm somewhere else. Rather, 'if you emit at one place, and also prevent an equal quantity of emissions at another place, you do no harm because you do not change the global concentration' (Ibid.).

This is true in one sense but not in another. It is true that if one causes an emissions event e_1 at time t_1, then prevents some emissions event of equal size, e_2, that would have occurred at some later time t_2, one makes it the case that the overall concentration of GHGs is no higher at t_2 than it would have been had one not caused e_1. But one *does* raise the

atmospheric concentration of GHGs at t_1. Now, on Broome's view as expressed in (Broome 2019), as we saw in the previous section, because of the butterfly effect, it is a statistical certainty that different emissions events of the same magnitude, despite having the same *ex ante* expectation of harm, will do different amounts of actual harm. The actual molecules of GHG produced in each case will interact with the atmosphere in specific, unpredictable ways. Arguably, therefore, Broome's dismissal of the argument against offsetting, on the grounds that it is analogous to attempting to compensate for doing harm by preventing some entirely separate harm, is inconsistent with his most recent characterisation of the causal relationship between emissions and harm.

Here is an analogy that is perhaps more fitting than Cripps's noise pollution case. Suppose I throw a stone up into the air in a crowded square. Clearly, my action has a significant positive expectation of harm, although it is not certain I will do harm, as it is possible the stone might fall to the ground without hitting anyone. Suppose that while the stone is in the air, I regret my hooliganism, and tackle my friend to the ground just as he is about to throw a stone exactly similar to mine. Before my stone hits the ground, I have successfully reduced the expectation of harm in the square due to the risk of being struck by a stone to the same level as it would have been if I had never thrown my stone (supposing my throw and my friend's were probabilistically independent events - he was not following my example). Nevertheless, it is perverse to say that because I prevented a second stone from being thrown, my initial act of stone-throwing was not wrong.

Perhaps Broome will say the claim that '[t]he harm done by emitting greenhouse gas is done only through the effect it has on the global concentration of greenhouse gas in the atmosphere' (Broome 2016 160) is not incompatible with the butterfly effect view. Again, there is a reading of the quoted statement on which this is true; unfortunately for Broome, it is not the one he needs for his dismissal of arguments against offsetting to go through.

Clearly, the harm caused by molecules of greenhouse gas is done through their concentration, in the sense of their presence in the atmosphere, their relative prevalence in comparison to molecules that do not interact with infrared radiation, like N_2 and O_2. But it is not compatible with the butterfly effect view to claim that GHGs do harm only by affecting 'one quantity' (ibid.), namely the measured global average concentration of GHGs in parts per million. If butterfly effects occur, they are by nature local, propagating outward from the very small to the very large, beginning from 'small disturbances at one time and one place' (Broome 2019 112). Going for a Sunday drive, Broome claims, will cause 'typhoons to form at quite different times and places' (Ibid. 113) than if one does not go for a drive. Thus, on the butterfly effect view, even if offsetting my emissions reduces net expected harm to zero, actual harm will be done to different people if e_1 but not e_2 occurs, as against if e_2 but not e_1 occurs. If, as Broome claims, the individual's duty to reduce her emissions is a direct duty, a duty owed to particular people, and Broome is right that equivalent emissions events at different times and places have wildly different effects, then Broome has not adequately dismissed the justice concerns about offsetting of the kind raised by Elizabeth Cripps.

Offsetting Via Negative Emissions Through Direct Air Capture

If the moral problem with offsetting is that it is tantamount to attempting to compensate for wrongdoing by paying someone else not to engage in wrongdoing, it might be thought that the least controversial form of offsetting from a justice perspective would be to take carbon dioxide out of the atmosphere, rather than preventing carbon dioxide from entering it. This form of offsetting – negative emissions through direct air capture – at least evades one criticism, namely that one cannot compensate for one's own wrongdoing by preventing someone from doing something they had no right to do anyway. Negative emissions through direct air capture would at least generate an expected benefit roughly equal and

opposite to the expected harm generated by the same quantity of emissions, even though, if we accept the implications of the butterfly effect view, their harms and benefits are to different people. For the defender of the claim that one has a duty to reduce the expectation of harm generated by one's emissions to zero through offsetting, generating negative emissions through direct air capture is arguably the surest mechanism.

The principal technology that is currently viable at scale for achieving this is the planting of trees. In an important sense, afforestation should indeed be regarded as a technological fix – afforestation schemes raise many of the practical, political and ethical concerns that have been associated with geoengineering more generally. They also present a unique set of concerns.

In order for afforestation schemes to achieve additionality (that is, to ensure they bring GHG concentration to a level lower than it would have been without the scheme) it must be guaranteed that they will keep carbon out of the atmosphere for at least the same amount of time that the carbon they are supposed to offset remains in it. According to the Working Group 1 contribution to IPCC AR5, '[t]he level of confidence on the side effects of CDR [Carbon Dioxide Removal] methods on carbon and other biogeochemical cycles is low' (IPCC 2013 469) – there are still significant gaps in the science behind using afforestation to generate negative emissions, including the extent to which afforestation contributes to positive forcing effects through changing surface albedo and ecosystem disruption. Forests typically take 10 years to reach their maximum sequestration rate, and after 20-100 years (depending on species) become 'saturated' and no longer produce net greenhouse gas removal (Royal Society 2018 24). About 15 to 40% of CO_2 emitted until 2100 will remain in the atmosphere longer than 1000 years (IPCC 2013 472), meaning that for offsetting to be effective, the carbon sequestered in tree biomass cannot be allowed to re-enter the

atmosphere.[23] The risk to effectiveness presented by fire, disease and resumed deforestation due to political and legal upheaval is therefore very great. This means that afforestation for offset can be seen as introducing a dangerous new threshold for harm: a stockpile of carbon that is at risk of being released at any time. In this respect, it gives rise to governance concerns similar to those raised with respect to forms of geoengineering generally considered more radical, such as Stratospheric Aerosol Injection: there is a danger of 'termination shock' (cf. e.g. McCusker et al. 2014) whereby failing to maintain the programme leads to a sudden dramatic spike in radiative forcing and attendant harms.

Say (as recently seems to have occurred in Brazil) a climate-denying political leader comes to power in a country in which forests have been planted for the purposes of offset, and allows those forests to be burned to clear land for cash crops. One of the individuals who originally purchased the offset could argue that it is the political leader, and not the offsetter, who is responsible for the emissions relating to this act. The offsetter's emissions (or a nominal proxy for them) were safely locked away in the biomass, and it was the political leader, not the offsetter, who carelessly released them. But clearly, if instead of offsetting, the individual offsetter, and all those others whose purchases of offset contributed to the planting of the forest in question, had refrained from emitting in the first place, a risk would never have been created. The group of offsetters can therefore be viewed as participating in a collective harm, insofar as they together give rise to this risk of a sudden spike in carbon emissions.

A final ethical problem with afforestation for offset can be seen to arise in the following way. There are limits to how much carbon can in practice be offset through afforestation.

[23] IPCC AR3 stated that CO_2 remains in the atmosphere for 5-200 years, but this estimate was removed in subsequent reports because, in the words of contributing author Richard Betts, 'the lifetime estimates cited in previous reports had been potentially misleading' (quoted in Inman 2008), given that around 20% of the CO_2 increase will remain in the atmosphere for many millennia, or 'essentially forever' (Archer 2010).

It is estimated that by 2100, it will be possible to sequester a maximum of between 4 and 12 $GtCO_2$ per annum through afforestation, globally, with the variance in the estimate due to differing estimates for land availability (Royal Society 2018 24). In 2017, the USA produced 5.14 $GtCO_2$.[24] Therefore, it is possible that the maximum capacity the planet has for carbon sequestration through afforestation is equal to the emissions of a single country. What this suggests is that Broome's assertion that the harms of individual emissions are easily avoided is dependent on the assumption that universal compliance with the duty to offset will never be achieved (and indeed, that we will never even approach full compliance). It looks highly counter-intuitive, in principle, to say that we all have a duty of justice to do something that it is only possible on the assumption most of us do not do it.[25] In effect, Broome's claim that it is 'easy' to reduce the net expected harm of one's emissions to zero counts against him: if is 'easy' to reduce one's emissions to zero through offsetting, then we should expect a high degree of compliance. But if we expect a high degree of compliance, then the expected value of my combined action of emitting and offsetting is negative, given the constraints on the availability of offsetting. If we refuse to calculate the expected value of my combined action on the assumption of a high degree of general compliance, then we are, in effect, advocating a general policy whose justification relies on the policy not being adopted. There therefore seems to be a sense in which any argument in favour of meeting a duty of justice by offsetting through negative emissions is self-defeating.

[24] US Energy Information Administration 2018, available at https://www.eia.gov/todayinenergy/detail.php?id=36953 (retrieved 07/07/19)

[25] Kai Spiekermann makes a similar point, observing 'claims that we can fly as much as we want (as long as we offset) or drive big cars (as long as we offset) are a lot less convincing when it becomes transparent that there are not enough offsetting opportunities to go around for everyone' (Spiekermann 2014 925)

Is There an Individual Direct Duty to Reduce Emissions to Net Zero?

Neither Broome, Lawford-Smith or Nefsky give us a convincing argument for the claim that there are direct duties incumbent upon individuals not to produce GHG emissions, grounded in a duty not to harm others. This means none of their views can be invoked as a solution to the paradox of collective harm: they cannot be used to ground our intuition that wrongdoing has occurred in the production of collective harm, and of climate change considered as a collective harm in particular.

As we saw, Broome's argument that individual contributions to collective harm are prohibited by direct duties failed, for three reasons. First, it is not clear that individuals directly cause actual harm at a constant rate in proportion to the quantity of emissions, meaning it is not clear that small quantities of emissions directly cause actual harm. Second, the claim that it is wrong to impose a risk of harm is under-supported – at least, it has not been shown that acts which produce small quantities of emissions are among the subset of risky acts that count as unjust. Finally, it is not clear that it is easy to reduce one's emissions to net zero, meaning individuals retain a defence against the charge that they act unjustly, on the basis of demandingness. Offsetting itself likely implicates individuals in further collective harms. If we retain the intuition that some wrongdoing must be going on when collective harm occurs, then there must be something wrong with offsetting. And indeed, Broome seems to have retreated from the original claim. He denies that individual contributions to GHG emissions make no difference: they make a difference in that they do expected harm. But he does not deny that there are cases in which actions which produce emissions are justifiable, and suggests that Sinnott-Armstrong's joyguzzling case may be just such a situation.

Julia Nefsky and Holly-Lawford Smith were both responding to a particular framing of the collective harm problem, the claim that individuals have no reason not to contribute to collective harms because they make no difference. Nefsky offered us a reason not to contribute to collective harms, on the grounds that we have a reason to be a non-superfluous part of the cause of some positive collective impact. Lawford-Smith went further, arguing that in contributing to GHG emissions, one might be wholly causally and morally responsible for some threshold-effect negative impact in the order of severity of causing 10 deaths. Where Nefsky's view proves too little, Lawford-Smith's proves too much. The idea that by contributing to some collective impact, I am helping to produce that impact, gives one a reason to contribute, but it is a reason that is easily outweighed, and on its most plausible rendering collapses into the view that I have a reason to contribute because doing so carries some expectation of benefit. The idea that individuals jointly trigger threshold effects, meanwhile, generated the highly counterintuitive conclusion that any contribution to GHG emissions might make an individual morally responsible for all the harms of climate change throughout the world, and extending far into the future.

Appeals to threshold effects, appeals to direct harm, and appeals to considerations of expected harm, are arguably the most significant and plausible ways of arguing that individuals have direct duties to mitigate their emissions grounded in the duty to avoid harming others. The fact we have shown these arguments to be at best highly problematic should give us strong reasons to prefer other ways of grounding individual responsibility for GHG emissions, if other such methods are available. In particular, it provides a strong argument for looking beyond considerations of what single individuals owe to each other, towards considerations of duties which arise from our relationship with certain kinds of groups. As we saw in Chapter 2, it plausible that because groups produce emissions of sufficient magnitude to cause significant climate change, moral fault should be assigned at

a group level. This would mean that groups bear responsibility for climate change, that remedial duties are incumbent upon groups. The challenges facing this kind of approach are a) specifying a kind of group that is capable of bearing responsibility, and b) explaining the relationship between group fault and individual duty. In the next chapter, we will examine some recent contributions to the literature that explain that individuals are responsible in relation to group-caused harms as a result of their membership of faulty groups. While not wholly successful, they will provide the outline of a more cogent account of the nature of individual responsibility for participation in collective harm.

4. Responsibility and Proto-Shared Agency

To reiterate, a 'responsibility gap' occurs in situations in which uncoordinated groups of individuals together cause serious harm, but no single agent in that group can be said to cause harm. We speak of a gap, because there seems to be a sense in which some agent or agents *ought* to be morally accountable for avoidable, serious, foreseeable, human-caused harms, and yet no appropriate candidates can apparently be found. When agents cause serious harm as part of a joint enterprise, no such gap occurs, as we can in such cases assign responsibility to the group itself. The question of how remedial responsibility is to be distributed may remain somewhat vexed, but the more fundamental philosophical problem does not arise: there is at least some entity that we can hold to account.

For this reason, an appealing approach when trying to bridge responsibility gaps is to posit some minimal degree of group coordination, so that the group can be considered an entity capable of bearing some form of collective responsibility, which correlates with determinate duties on the part of individuals. Appeals to collective responsibility, however, have to tread a fine line: posit too much coordination, and we no longer have a realistic description of the relations between actors in key problem cases. Climate change (which forms my core example of a responsibility gap) is a particularly thorny problem precisely because it is caused by individuals widely dispersed across time and space, making significant group coordination an implausible scenario.

To be clear, we can – and do – assign *causal* responsibility to uncoordinated groups. There is nothing controversial about the claim that a group of polluters is causally responsible for the total pollution they together create. Moral accountability, however, seems to require agency, to the extent that only agents are capable of bearing remedial duties. Remedial

duties are duties to perform some action, and non-agents cannot act. Nor does it make sense to talk of individuals having 'fair shares' of remedial duties that would be borne by an uncoordinated group, were it a group agent. Just because some group of 100 people, were it a group agent, would have the duty to remediate some quantity of harm, it does not follow that each member of that group has a duty to remediate one hundredth-share of the harm. There is no duty to be distributed because the duty does not fall upon the non-agent group in the first place.

In the literature on group agency, two towering figures are Margaret Gilbert and Michael Bratman. Gilbert views collective agency as arising from 'joint commitment'; Bratman views it as arising from first-personal planning norms, the norms that apply if I intend that *we* ϕ. Neither of these phenomena can obviously be said to apply in a case such as climate change, at least not to a sufficient degree wholly to close the responsibility gap. It does not appear that contributors to GHG emissions have given one another any kind of manifest commitment that they will play their part in performing this activity (see Gilbert, e.g. 2013; 2009), nor is it obvious that they 'mesh subplans' to carry out some shared intention (see Bratman, e.g. 2014; 1993). Nevertheless, it has been suggested that more minimal forms of group coordination, approximating those described by either Gilbert or Bratman, might be said to come into play in the kinds of collective impact cases which give rise to a responsibility gap. Notably, Elizabeth Cripps (2013) has suggested that such sets of agents may constitute a 'collectivity', defined as group coordinated through their shared *interests* (Cripps 2013). Although Cripps sets out her view partly as a response to Gilbert, her approach can arguably be viewed as a pared-down version of a Bratman-style account, as no role is given in her account to any kind of performance of commitment, and shared interests can be viewed as being something like implicit or potential shared participatory intentions. More recently, Stephanie Collins has argued that individuals in such groups

might have 'exchanged commitments' to pursue particular goals, and that this fact might ground duties to remediate harms generated in the pursuit of those goals (Collins 2017). This can be viewed as a more minimal version of a Gilbert-style account, as the kind of 'exchanged commitments' posited are not intended to be commitments to perform some action *together* but rather simply a mutual manifestation to the other party *individually* to 'positively respond to some permissible end' (Collins 2017 585).

In order to be successful, these accounts need to fulfil two conditions: i) the degree of coordination they describe must be sufficiently *extensive* to ground the kinds of duties these theorists take to arise from them, and ii) they must be sufficiently *minimal* accurately to capture the relations between agents in central examples of the collective impact problem. In what follows I will argue that these authors fail to meet these conditions simultaneously. Finally, I will give a sketch of an account that does better on the measure of these conditions, building on the work of Christopher Kutz. I will be chiefly concerned to defend the view from the charge that it is unrealistic.

Exchange of Commitments

Collins summarises her account as follows:

> *If: (i) two or more individuals have exchanged commitments to one another to positively respond to a permissible end, and (ii) harm arises from any (including aggregations) of those individuals' reasonable positive responses to that end (including the responses of realising, pursuing, endorsing, maintaining), and (iii) individual duties to remedy the harm cannot be justified on the basis of individual harms or wrongs, then: (iv) each of the individuals has a duty (owed in part to those with whom she exchanged commitments) to take on costs in remedying the harm.* (Collins 2017 585)

Collins's central illustration (a case adapted from Miller 2011) concerns two musicians, Bert and Charles, who each exchange a commitment to give free concerts in the town square. This commitment is supposed to be something like a promise, but not necessarily 'as explicit as promises typically are' (Collins 2017 587). Following Gilbert's conception of joint commitment, exchanged commitment is supposed to be something that can be manifested through positively affirmative behaviour, or even through inaction in certain institutional contexts. This is supposed to extend the number of situations in which an exchange of commitments can be said to have occurred, so that it can be said to provide a description of a wider range of real-world phenomena. Unlike Gilbert's notion of joint commitment, however, we are not supposed to think of the two musicians as engaged in some joint venture, for example of ensuring people in the square are entertained every evening of the week. Rather, both musicians simply manifest to each other their intention to perform in the square regularly.

During one concert, Bert accidently runs over Anne's foot with his piano. This fact, let us assume for argument's sake, renders Bert individually liable for Anne's medical bills. But Bert has no money. Who should pay Anne's bill? Collins's suggestion is that Charles has some duty to pay for Anne's medical bills (though perhaps not full liability), which 'arises out of his and Bert's exchanged commitments to the end, which generates a duty to support one another in the reasonable pursuit of the end' (Collins 2017 589). Assuming Charles and Bert do not constitute a group agent, Charles's duty does not arise directly from Anne's claim to compensation, but from Bert's claim to support from Charles in their common (although not *joint*) project.

Individuals united by exchanged commitments, on Collins's view, constitute 'a weak type of proto-shared agency, deriving from their common aim, common dispositions to predict, rely upon, and reinforce each other's actions, and common rational availability of we-

reasoning decision-making processes' (Collins 2017 589). She points out that, of the groups that meet the conditions for proto-shared agency, many also meet the conditions for shared agency proper. Depending on the nature of Charles and Bert's agreement, they may in fact be a group agent. Thus, Collins needs to be careful if she is going to convince us that, if gap-filling remedial duties do arise in such cases, they arise from the group's status as a proto-group agent and not from their status as a full group agent. If Bert and Charles are involved in a joint enterprise – they are two members of a band putting on a concert together – then it is more plausible that Charles should be on the hook for Bert's mistake. But it is not easy to see why exchanging promises individually to pursue the same aim as some other person should make one accountable to that other for support in his pursuit of that goal, unless that is expressly part of the promissory arrangement.

Collins's view is that a claim to support is the 'flipside' of a relationship of accountability. If you and I have exchanged commitments individually to cut our carbon emissions by a certain amount this year, I have standing to hold you accountable (to chastise you in some way) if you fail to meet your commitment. By the same token, the thought goes, if you are struggling to meet your commitment, I have some duty to provide you with support, for the reason that if I do not, 'you can question whether my commitment to emissions reduction was genuine in the first place' (Collins 2017 587), and hold *me* to account on that basis. In Collins's examples, if you are struggling to stick to your vegetarianism, I might have a duty to offer consolation and advice, or if you are unable to keep cycling to work because of a broken bicycle, I might have a duty to help you repair it. These are perhaps not the most persuasive examples, as you might think such behaviour was mandated by common decency alone. Nevertheless, we can agree there is some force to the idea: it is certainly the case that fellow travellers in pursuit of some common cause they greatly value might feel special duties to one another.

A suspicion against the account remains, however, as it is very difficult to determine, in such cases, when a common cause becomes a joint cause, when a common aim becomes a shared intention. It may be that our judgement about the duties of the two environmentalists arises from our viewing them has having committed to a project of reducing their *combined* carbon emissions by a certain amount, rather than committing each to reduce their emissions by a certain amount. We could control for this potential "noise" in the following way: we might imagine that the two are not friends, but rivals, and want to outdo each other in their commitment to emissions reductions. They exchange promises not in order to cooperate, but so the other can monitor their progress.[26] In this case, it looks strange to say that one should have a duty to help the other, because it was obvious from the start that they never took themselves to be taking on any such commitments. As is common in the group agency literature, for example the exchanges between Gilbert and Bratman, it is open to dispute whether this updated example constitutes a core case of commitment exchange, or whether added conditions are being covertly brought in – an implicit waiver of the right to support. While I don't think I am able to settle this point, the updated case gives us prima facie reason to doubt that Collins has identified a phenomenon different in kind from shared agency, which gives rise to remedial duties.

[26] I take my cue here from an example from Anna Stilz (2009 179), which is intended to refute the view that 'strategic coordination' is sufficient for collective action (see Lewis 1969; Hardin 1982). The case concerns two competitive society ladies, Mrs. Pennypacker and Mrs. Vandalay. Mrs. Pennypacker loathes Mrs. Vandalay, but nevertheless intends to attend any party attended by her rival, in order to ensure she is regarded as "queen bee". Mrs. Vandalay has corresponding intentions with respect to Mrs. Pennypacker. These ladies coordinate their actions as meticulously as they would if they were cooperating, but in fact they are working against each other. We could easily imagine that these ladies have "exchanged commitments", in Collins's sense, to act as they do; it is even plausible that one lady's response to the other's failure to attend a party could take the form of a reproach (though tinged with vindictive glee). Yet it is highly implausible that either has any duty to support the other.

There are a number of features of the view that remain to be clarified. For one, we may wonder how far exactly a duty of 'support' can be said to go. Collins's acknowledges that Charles may not be 'on the hook' for the whole of Bert's liability to Anna. We might think we are therefore owed some account of what should count as doing enough to discharge one's duty. Could Charles fulfil his duty of support in other ways, for example by showing sympathy for Bert's predicament? On a related note, it is unclear to what extent the normativity of remedial duties is generated by commitment to the common project, and to what extent it is generated by the promissory relationship between the two agents. The following question brings this ambiguity into sharp relief: if I could do more for the cause by not supporting the person with whom I had exchanged commitments than I could by supporting her, what would my primary duty be? Say that in the case of the mutual commitment to emissions reductions, the agents did not set any specific target, but simply committed to reducing their emissions as much as they could. Perhaps, when asked to help repair my friend's bicycle, I am on my way to a meeting for an important investment opportunity in green technology, which will reduce my carbon footprint significantly more than the amount my friend's will be increased by having to drive to work. Suppose further than I am already on track for significant reductions, so that it is not the case that, were I to miss this opportunity, I could be said to be failing to meet my personal commitments. If we judge that I ought to help my friend, this would indicate that the normativity is generated by the nature of my relationship with the friend; if we judge I ought to attend the meeting, the normativity seems to be primarily generated by the force of my first-personal planning commitments.

Importantly, though, this ambiguity is arguably not innocent: it masks an inconsistency in the application of the notion of support in the musicians case as compared with the emission reduction case. I have an obligation to support my friend in her plan to cut

emissions, because, in part, of my own commitment to cut emissions. It would be inconsistent of me not to help her, given my own professed aims. Bert's injuring Anne, and the duty of compensation that arises from it, however, is orthogonal to the common project of playing music. It is not clear why Charles's paying compensation to Anne should count as supporting Bert in his commitment *to play music in the square*. It looks doubtful that Charles's refusal to cover Bert's liabilities would count as evidence of a lack of commitment on Charles's part to their common aim. It might be countered that perhaps covering Bert's liabilities *is* necessary to allow him to keep playing – perhaps allowing him to become embroiled in a tortuous court battle would frustrate his musical project, or perhaps he will be banned from the square unless he makes peace with her. But this kind of modification would not be available in all the cases to which Collins wants the model to apply. The hope is that this model of proto-shared agency can go a significant way towards bridging the responsibility gap in the case of climate change, because even if no individual can be held accountable for their personal contributions to GHG emissions, most individuals are involved in structures of commitment exchange at a larger group level, which would leave them 'on the hook' for the macroscopic emissions-related harms that can be said to be perpetrated by the group. Many people, the thought goes, can be said to have exchanged a commitment with their colleagues to 'keep this company in profit for the next few years', or some other similar end. Collins believes the upshot of this is that they can be held accountable, to at least some degree, for the harms perpetrated by that 'end-oriented group', the company. But it is not the case that a refusal to accept responsibility for harms perpetrated by the company can be used by my colleagues as a sign of my lack of commitment to the project of "keeping this company in profit for the next few years". Remember that it should not be shared agency that is doing the work here, but proto-shared agency. This means anyone who has exchanged commitment to the end of keeping

the company in profit, such as consultants and contractors, would have to be included in the end-oriented group. It is implausible to suppose that the bare commitment to keeping the company in profit, considered in isolation from any shared intention or shared agency the employees of the company might have, necessarily implies that the employee takes on liabilities with respect to damages the company does as a group.

"Collectivities" and "Should-be Collectivities"

Both Bratman's and Gilbert's accounts of collective agency require some form of shared intention. For Gilbert, this shared intention exists at the level of what she calls the 'plural subject', a kind of implicit entity that makes it the case that it makes sense to say "the group intends to ϕ", independently of the intentions of its members. For Bratman, shared intention subsists purely at the individual level: we have a shared intention to ϕ when: (i) I intend that we ϕ, and (ii) you intend that we ϕ, and I intend that we ϕ in accordance with and because of (i) and (ii), meshing our subplans, and you do too, and all this is common knowledge between us (See Bratman 2014, 40-59; 1993). As already intimated, however, shared intention in either of these senses is implausible in large-scale groups, widely dispersed across the world, with limited communication between members. For this reason, Cripps suggests that there is a notion of 'collectivity' that can do important normative work while moving away from the 'intentionalist model' (Cripps 2013), the view which sees shared intention as necessary for having the kind of status that makes groups capable of bearing duties.

On Cripps's account, it is not necessary that participants in a "collectivity" are internally coordinated. On Gilbert's view, collective agency requires some kind of individual manifestation of joint commitment, which may be anything from a contract or an oath of allegiance to mere body language, depending on context. For Cripps, meanwhile, members of a collectivity need not be coordinated in any way; indeed, individuals need not even be

aware of each other's existence. All that is required is that individuals are dependent upon others for the achievement of their goals, or for securing their 'fundamental interests'. Thus, for example, she claims that a group of castaways washed up on different parts of a small desert island, although unaware of each other's presence, may constitute a "collectivity" because of their dependence on one another for their survival.

She defines 'collectivity' as follows:

> *A set of individuals constitutes a collectivity if and only if those individuals are mutually dependent for the achievement or satisfaction of some common or shared purpose, goal or fundamental interest, whether or not they acknowledge it themselves.*
>
> (Cripps 2013 28)

We may note therefore, that this view differs from both Bratman and Gilbert's views in two ways. It requires group members to be mutually dependent on one another, and it stipulates that this dependence may be in relation to an *interest* rather than the achievement of an intention or joint commitment. It is thus in once sense more restrictive than their conceptions and in another sense more expansive. The mutual dependence condition may look a little surprising: it would seem to rule out several cases of shared agency that Gilbert and Bratman would count as core cases. Bratman's key case of two people painting a house together, for example, is not obviously one in which either painter is dependent on the other: it may be that each is perfectly capable of painting the house alone, but they can simply do it faster if they work together. Cripps clarifies this point: a group can be mutually dependent for the satisfaction of some end when it is one that 'if achieved for any, must be achieved for all' (Cripps 2017 33). Thus, in her example, Fathers for Justice is a mutually

dependent group in the sense that, if their aim of greater recognition for the rights of fathers during divorce proceedings is achieved, it will benefit all those who share that interest.

Cripps thinks the notion of collectivity needs to be more expansive because of the existence of groups she thinks are clearly "collectivities" in some sense, but which are ruled out by 'intentionalist' views. A key case is families: babies, she thinks, are clearly members of family "collectivities" although they are unable to participate in shared intention because they are unable to manifest joint commitment, or lack the conceptual resources to have "we-intentions". Rebellious teenagers may fume "I don't want to be part of this family anymore!", but nevertheless take advantage of household amenities. What makes sense of these cases, Cripps believes, is that these children remain in a relationship of mutual dependence for the achievement of fundamental interests along with their other family members, and thus remain a member of the family collectivity.

In order to determine whether Cripps is right to assert that "collectivity" needs to be understood in a way that includes babies and rebellious teenagers, we need a clearer idea of what exactly a collectivity is, what explanatory function it is supposed to serve. Remember that Gilbert's view begins as an account of collective action. In setting out the view, Gilbert introduces the concept of the 'plural subject', and this concept is then used in describing other group-coordinated states and attitudes, most notably collective belief. The plural subject, on Gilbert's view, is constituted by the various beliefs and intentions that the group has allowed to stand as its joint determinations, often through some institutional procedure such as voting. This picture of collective intention forms the foundation of collective agency: if I am acting according to our joint commitments, we are engaged in shared agency, whether or not I privately intend or approve of the whole project. This analysis thus helps to describe the phenomenon of political obligation, whereby we are committed to obey the state even though we do not agree with all of its determinations. Cripps, meanwhile, refers

to the object of her analysis only as a "collectivity", apparently without stipulating that this notion should necessarily imply collective agency. Thus it is arguable that she is able to eliminate the requirement for some kind of shared intention from the view simply by ceasing to describe a kind of agency. If this is right, then it is somewhat misleading to position her view as an alternative to Gilbert's, as it looks like she is describing a different kind of phenomenon.

What then is the functional role of a "collectivity"? As Cripps characterises the concept, it has three features: i) the significance of certain acts by individuals cannot be described without reference to the whole, ii) what the collectivity does is distinct from what the individuals do, iii) collectivities persist even as individual members are replaced. Thus the concept of a collectivity apparently plays a kind of descriptive-explanatory function. Let us return to the case of the mutually dependent castaways. Suppose the castaways are not just unknown to each other, but are in fact enemies – rival gangs – and as a result fail to cooperate. In a case such as this, it is especially difficult to understand in what sense the group itself *does* something. Cripps points out that nevertheless there are descriptions of their behaviour that are not reducible to individual action: 'the castaways destroyed themselves', for example. No particular castaway destroyed himself: this description only makes sense if we regard it as applying to the castaways as a collective entity. We should note, however, that this is still not to say that the castaways exercised any kind of agency, any more than the sentence 'the forest teems with life' implies that the forest performs the action of 'teeming'. There are just certain things individuals do, or in this case fail to do, which make it the case that the group can be described as 'destroying itself'.

It is evident from Cripps's work, including earlier work (see Cripps 2011), that Cripps is using the term 'collectivity' to mean a group that plays an irreducible role in the explanation of social phenomena; she cites Paul Sheehy (2006) as a key influence on her approach.

Her conception of 'collectivity' may in this way be viewed as a contribution to the debate between methodological individualism and holism in the philosophy of social science. If this is right, it does seem inappropriate for her to treat her view as a rival to those of Gilbert and Kutz, who are concerned with the conditions of shared agency. Cripps, then, is justified in her view that shared intention is not a necessary condition of 'collectivity'. But it is nevertheless the case that so called 'intentionalists' have no reason to be threatened by her view, or so I am claiming.

Whatever the genealogy of Cripps's concept of collectivity, at the very least, for Cripps's argument to go through, collectivities need to be relationships through which 'we acquire special duties'. Most significantly, she claims, 'a set of human beings (moral agents) who are mutually dependent through a common fundamental interest have a weakly collective duty to secure that interest' (Cripps 2013 48). Cripps suggests that a group of second-home owners whose properties surround a village constitute a collectivity, insofar as they have a mutually common interest in a well-maintained green, and that this is true even though these people may visit the village rarely, have never met, and have no idea whether any of the other residents uses the green. If the interest is sufficiently important, they would also, on Cripps's view, have a weakly collective duty to maintain the green, grounded in a 'collectivized weak principle of beneficence' – the thought being that if, through collective action, one can participate in the production of benefit for others at little cost to ourself, then one ought to do so.

Some might argue this claim makes light of persistent collective action problems, offering too easy a resolution. Such problems are characterised by a 'problem of coordination', insofar as '[w]hat any member of [the group] ought to do depends on what others do' (Kutz 2002 476). Distributive duties to maintain the green only hold in contexts in which the individual can rely upon a sufficient number of others to participate. If they held without

this caveat, it would be wrong to expect compliant individuals unilaterally to incur costs when to do so would achieve nothing. For this reason, it might be argued, Cripps's claim that shared participatory intention is not required for shared remedial duty looks problematic.

Cripps, to be clear, does not claim that individuals have a duty to act as though they were playing their part in a cooperative scheme no matter what other members of the group are doing. Such duties, which she calls 'mimicking duties' (mimicking one's role in a cooperative scheme that does not yet exist), are of at best secondary importance in comparison to 'promotional duties' – the duty to promote the establishment of a cooperative scheme, on her account. Considerations of demandingness are therefore less of a problem on her view then they are on the kinds of views Kutz was rejecting. Plausibly, it is not too demanding to recognise a duty to promote cooperation as best as one can, even in a deeply uncooperative society. We can view this duty as inherently dynamic in its content, its degree of demandingness increasing in proportion to the degree of responsiveness of other members of the collectivity. If this is right, though, then Cripps must acknowledge that a weakly collective duty grounded in moralised collective self-interest would at a limit be very weak indeed: essentially just a duty to be disposed towards cooperation, should the opportunity present itself (if so, then it parallels a view defended by Felix Pinkert (2015)). This would mean that, like the accounts of Feinberg and Held referred to in Chapter 2, Cripps's concept of weakly collective duty grounded in moralised collective self-interest would essentially be a way of locating distributive fault for group-level failures. Given a sufficiently uncooperative social environment, it would not necessarily provide us with the sought-after means of reading off remedial duties.

Cripps, moreover, wants to claim not only that collectivities bear remedial duties, but that an even more loosely associated kind of group, 'potential collectivities' or 'should-be

collectivities', bear such duties. These are groups that do not even have a common interest but are 'a prima-facie locus for moral condemnation' (Cripps 2013 68). Thus the group of 'polluters', as an uncoordinated group causally responsible for the harms of climate change, which constitute a threat to fundamental interests, can be regarded as a group that "should" have a common interest or shared aim, namely the aim of remedying harm, and thus has a 'weakly collective duty' to collectivise in order that it is capable of carrying out that duty. To illustrate this point, Cripps offers the case of a group of swimmers who, in an uncoordinated way, are together causing so much turbulence in the water that it is causing someone to drown. These swimmers, providing that they could reasonably be expected to have foreseen that their combined actions would cause harm and that their contributory action was avoidable, should, Cripps argues, be regarded as the bearers of 'weakly collective responsibility', which is said to give rise to a duty to organise themselves so as to prevent such harm, via a 'collectivized no harm principle'.

Remember, however, that one of the premises that motivates our entire discussion is that assignments of collective responsibility are difficult in the case of climate change because the group of polluters is so widely dispersed across time and space. There are significant differences between the group of swimmers and the group of polluters. For one thing, in the case of the swimmers, their ability to cooperate, and the sure effectiveness of their cooperation, is plausibly common knowledge between them. It is obvious from basic assumptions about human behaviour that it is within their power all to stop thrashing about. All that it then takes is for certain swimmers to manifest their intention to stop swimming. Thus the conditions for Bratman-style collective agency are firmly in place. In the climate change case, however, it is far less easy for temporally and spatially dispersed individuals to manifest their participatory intention, and far less easy for them to have the same confidence of success. Thus the problem of coordination problem remains a serious

obstacle: it is arguably unreasonable to assign costly duties to promote collective action, when they have no means of determining whether compliance is likely at a sufficiently high level to be effective. If these duties are not to be regarded as especially demanding, then we have to question how far we have progressed with the task of filling the collective duty gap.

Quasi-Participatory Intention

It thus appears that, for these accounts to effectively close the responsibility gap, Collins and Cripps implicitly rely on the presence of a greater level of coordination than can be accommodated if we are to provide a realistic description of the climate change case. Collins's account of mutual commitment, it would appear, relies on the idea that I should be 'on the hook' for harms perpetrated by uncoordinated goal-oriented groups of which I am a member. This claim looks under-motivated, except in those cases in which such goal-oriented groups also participate in more explicitly coordinated forms of shared agency - such as corporations. Cripps's account purports to do away with the requirement that groups capable of bearing collective duties must share intention, but it is arguable that we lack an iron-clad case of a group of individuals who do not shared intention and nevertheless have clear participatory duties, beyond the duty to hold a cooperative disposition . Her village green case does not obviously give rise to any collective duties when the group of second home owners lacks shared participatory intentions, and her swimmers case seems to rely upon the ease with which swimmers would in such a case be able to coordinate shared participatory intentions, without the interpolation of Gilbert-style structures of join commitment. It thus looks implausible that a globally diffuse group – polluters – can be considered accountable for a failure to collectivise.

As Christopher Kutz has argued, from the perspective of a victim of global climate change, it is reasonable to regard the group of contributors as collectively responsible for a collective

harm. The problem, as Kutz sees it, is to square common-sense moral motivation from the perspective of the victims, with the perspective of the contributors, in order to demonstrate to contributors that it is indeed reasonable to regard *themselves* as on the hook for their involvement in collective harm. Cripps's and Collins's views can be read as attempts to carry out this project: they both claim that involvement in end-oriented groups should be enough to generate certain kinds of moral motivation, either through a sense of shared commitment to the end in question, or through a recognition of mutual interest and interdependence. Unfortunately, as we have seen, neither of these structures of motivational inter-connectedness with other agents seems sufficient to ground recognition of reparative responsibility. But this is not to say that they are not going about it the right way. Kutz suggested similar strategies for inculcating the necessary kind of moral motivation. Indeed, he seems to suggest both Cripps-style interdependence and Collins-style common commitment as potential avenues for generating motivation. Perplexingly, they are introduced not as separate ideas, but as different facets of the same approach. It is not immediately obvious how the two ideas are connected. In what follows I wish to argue these two ideas can indeed be regarded as a single approach, and that this approach can form the basis of an adequate case for the kind of moral motivation for which we are searching. That is to say, I will describe a form of proto-shared agency that finally allows us to bridge the responsibility gap.

Kutz writes 'one part of the task of dealing with collective harms is emphasising the moral significance of pre-existing networks of moral motivation' in the form of 'overlapping fields of shared meanings and political identifications' (Kutz 2000, 189). Thus, 'American drivers' can be morally motivated to reduce their emissions because 'thinking of the damage that I and my fellow American drivers do confirms me in a regional identity I already hold

(Ibid.).[27] The point is introduced, however, in a context in which interdependency is placed in the foreground. Referencing Bourdieu, Kutz notes the dialectical relationship between our shared values and the political circumstances in which we make choices – this itself, he suggests, can motivate an acknowledgement of accountability for the social consequences of those choices. The values of 'comfort and privacy' reflected by driving are only possible given social conditions of 'cheap fuel and disguised public subsidies' for the oil and automotive industries (Kutz 2000 188). The presence of these political conditions in turn reflects a valuation of privacy and comfort in the individual agent.

In later work Kutz develops this latter point about interconnectedness, apparently letting the earlier focus on political identity fall into the background. Greenhouse gas emissions from large cars, it is argued, can be viewed as the result of a joint enterprise, because individuals' choices to buy large cars are conditioned by similar choices by other consumers – consumers have a socially conditioned sense that SUVs are safer, more stylish, or indicative of social status. This interconnectedness between individuals' choices constitutes a certain 'zeitgeist', meaning 'we can say that the global increase of CO_2 emissions is attributable to the collective, and not merely parallel, acts of US consumers' (Kutz 2015, 360). The sense that we are as consumers engaged in a joint enterprise can engender a useful sense of 'collective guilt', that can 'induce the participation of individuals in schemes to solve collective harms' (Kutz 2015, 354).

This seems rather quick. Why exactly should the recognition that our collectively harmful choices are conditioned by similar choices on the part of others make a guilty reactive emotion more appropriate? As Iris Marion Young has argued, we might think the

[27] Kutz makes the point using the example of contributions to the emission of chlorofluorocarbons in relation to the degradation of the ozone layer, but the point holds equally for contributions to the greenhouse effect

converse: that the fact my choices arise from patterns of behaviour considered socially normal prevents them from being the proper objects of guilt or blame (Young 2011 103). The realisation that my choices are socially conditioned may lead me to view them as less than fully mine, and thus to a denial, rather than an acceptance, of accountability.

Arguably, the thought is something like this: purchasing a large car because of the prestige one hopes it will bring is something like joining a club. Even though there is no real coordination between the purchaser and the existing community of drivers, one shares with them a kind of minimal participatory intention: the intention to be members of the group of SUV drivers (together). There are, however, certain obstacles to regarding this quasi-participatory intention as grounding recognition of accountability. For one, it does not seem correct to regard this as an intention that *we* do so-and-so. Thus, not only it does it not seem correct to regard the action as being something that we do together (which would ground the attribution of complicitous accountability), it does not even seem plausible that I *think of myself* as acting as part of such a *we*. It is not necessarily my intention that we, the group of SUV drivers, drive SUVs. In fact, I might prefer it that as few other people as possible are able to buy SUVs, as this will increase my relative prestige.

Kutz has his own account of joint action, one that is more individualistically reductive than (for example) Gilbert's. On this view, joint action does not require any special kind of shared intention (of a plural subject), rather, it requires a particular kind of individual intention: participatory intention. On Kutz's account of participatory intention, to have a participatory intention is to intend 'to participate in a collective act' (Kutz 2000 73). To intend to perform some action (following Davidson) is to behave in a way that is causally and teleologically explained by the agent's goals (Kutz 2000 72; Davidson 1980a, 1980b). To intend to participate in a collective act, therefore, is to behave in a way that is explained by having as one's goal that the group does something together. The question, then, is why

I should view driving SUVs as a collective act. Perhaps it will be argued that in order to "buy in" to SUV driving as a mark of social status, I must intend that other high-status individuals continue to drive SUVs, so that doing so continues to be prestigious and I remain able, through this action, to achieve my aims. Arguably, there is some minimum number of high-status SUV drivers that I need to continue driving their SUVs in order for me to achieve the goal of sharing in their prestige. Even if we accept this, however, there is no guarantee that the number of such drivers is large enough to leave me on the hook for significant harm. I may consider it enough that two or three individuals in my immediate neighbourhood continue to maintain the prestige of the vehicles; this is would hardly be enough to ground a significant degree of accountability for climate change.

Again, however, a response might be available on behalf of Kutz – this objection arguably fails once we have a more complete understanding of Kutz's suggestive appeal to 'zeitgeist'. When I purchase an SUV in order to gain prestige, it might be thought, it is part of the content of my intention that the entire system of values whereby SUV ownership contributes to social status endures. Were it not for this system, I would not be able to achieve my goal, so the continuance of that evaluative culture must, arguably, be implicitly included in my intention. This understanding of the view speaks to the remarks Kutz makes about the reciprocal relationship between choices and socio-political conditions, in connection to Bourdieu's concept of *habitus*. In choosing to purchase an SUV, it might be thought, one shares in the project of reproducing the conditions that made that choice seem appealing in the first place. One intends not only to partake in, but to contribute to, the continued prestige of SUVs.

It is not immediately clear how the conception of quasi-participatory intention just outlined squares with the initial appeal to the idea of *political identity,* drawing on the example of *American* drivers. A sense of identity, and participatory intention, seem to generate self-

directed reactive attitudes of accountability in very different ways. Accountability in relation to political identity seems to be an idea close to what Kutz has referred to as a 'moral...regimes of strict or vicarious liability' (Kutz 2015 348), by analogy with the concept of strict liability under the law. Just as one's identification with a family member motivates one to accept liability for the harmful consequences of their actions, so too does identification with one's country engender feelings of guilt for the harmful consequences the actions of one's fellow citizens.[28] If my child breaks a vase, I feel a duty to pay for its repair because if I decline such a duty, I thereby repudiate my child's relationship with me. Similarly, it seems Kutz wishes to say, accepting accountability for the damage done by one's fellow American drivers might confirm one in one's identity as an American (conversely, a refusal to accept accountability would be a repudiation of that identity).

Kutz himself notes the problem with this approach – it is not necessarily the case that solidarity with our fellow American drivers in the face of moral criticism would generate a guilty response. Rather, we are likely to respond in with the following sentiment: '[w]e better band together to protect our shared way of life' (Kutz 2000 187). If we identify in particular with the group of American *drivers*, then the emotional dynamic Kutz describes leads to a kind of impasse in the specific context of accountability for climate change: though we may be prepared to accept accountability for the actions of fellow American drivers, we would not be able to accept reparative consequences in relation to that accountability which undermine our continued ability to identify as drivers. Accepting responsibilities in relation to climate change mitigation, it might be thought, would do just that.

[28] This is a more general dynamic of accountability than *political responsibility*, in which individual accountability is grounded in the idea that one is a member of a collectively self-determining political community (see Young 2011 84, Arendt 1987 43-50).

Meir Dan-Cohen (whom Kutz cites in order to distinguish his own view from Dan-Cohen's) holds that the question of whether one is responsible when one's child breaks a vase is best understood as a question of where one draws one's 'boundaries as a subject' (Dan-Cohen 1992 963). '[B]y assuming responsibility for an object or event', Dan-Cohen argues, 'I also implicitly affirm a certain aspect of myself as a viable source of my authorship' (Ibid.), as when I accept that a car accident was the result of *my own* negligence. What Dan-Cohen calls 'object responsibility' – responsibility in the forward-looking sense over certain objects, or events, or persons, or aspects of character – provides the basis for 'subject responsibility' – agential accountability. Accepting that negligence is *my problem* provides the ground for my accountability, just as accepting a child is *mine* provides the ground of moral liability for damages it causes. Thus, on this account, when one accepts accountability for the harms caused by one's fellow American drivers, it is because one includes them, to some extent, in one's 'boundaries as a subject'. What we have been calling 'identification' with the group of American drivers would be cashed out as a kind of projection of the self so as to partially include them, a process of 'self-constitution' (Ibid. 966).

Kutz however responds to Dan-Cohen in the following way: 'rather than taking the collective harm or wrong to be an aspect of the agent's self, I prefer to understand it as a consequence of a shared venture with which one identifies' (Kutz 2000 187 footnote). 'Identification', it is being suggested, is just a kind of participatory intention, rather than as a projection of one's sense of self. Kutz leaves this puzzling equation of identity and shared participatory intention as a passing remark; it certainly requires further development if we are to make sense of it. What is the content of my intention when I identify as an American driver? Informed by the discussion of the case of SUV drivers, we must assume it is something like an intention to partake in and to maintain the culture, viewed as the American way of life, in which the freedom to drive is highly prized. This understanding

gives us a clearer view of the structure of the worry that people will 'band together' to protect their 'shared way of life' – the intention to partake in and maintain American driving culture is incompatible with accepting accountability for the damage done by the group of American drivers, since one could not intend to make restitution for one's involvement in that group, unless one lacked the intention to maintain American driving culture.

The task, then, is to identity participatory structures that would not require the agent to hold incompatible intentions if they were to effectively motivate self-reactive attitudes of accountability. The quasi-participatory intention involved in driving SUVs is more effective than the American drivers case, because one can intend to take on reparative duties for the actions of one's fellow SUV drivers alongside the intention to partake in and maintain the culture of prestige surrounding SUVs. One can simply recognise that one's intention has been misguided. The "American way of life", however, is a quasi-collective activity so imbued with affective significance that one would be unwilling to abandon one's participation in it. The case of SUVs driving is not special: there are many other such potential quasi-collective activities with respect to which participatory intention could ground a case for accountability. Identifying a sufficient number of them will allow us to describe a sort of patchwork of structures of accountability, which together will be sufficient to close the responsibility gap. Examples might be: the culture of cheap flights, the culture of expecting the same range of food produce to be available in supermarkets irrespective of season, which relies on global supply chains, the culture whereby people expect to buy new clothes frequently rather than repairing them.

Kutz claims that the quasi-participatory view is not a sufficient ground for individual accountability for collective harms unless it is combined with another source of accountability, which he calls 'symbolic accountability' (Kutz 2000 186). This latter term designates the sense in which individuals can be regarded as accountable 'in virtue of who

they are' (Ibid. 190). It is not obvious why Kutz does not believe quasi-participatory accountability is insufficient without symbolic accountability. A clue, however, can be found in the weakness of 'symbolic accountability' taken on its own: it is vulnerable to some of the same criticisms we saw levelled against virtue ethics approaches in Chapter 2. The point seems to be that individuals may be judged to have faulty characters because they participate in practices that *might* cause harm, and that by accepting the benefits of these practices they fail adequately to show concern for the victims. But in order for us to take performing the kinds of behaviours that produce collective harms to be a sign of bad character, we must first have sufficient grounds to link what the individual does to the harmful outcome. The quasi-participatory view can be thought of as providing exactly this link: it creates the idea of these collective harmful behaviours *as* shared practices, so that we can then recognise and regret the harmfulness of the practices at a group level. On the most cogent rendering of the view, then, quasi-participatory accountability and symbolic accountability are not two separate sources of norms. Rather, 'symbolic accountability' denotes the fact we can be reproached for faulty conduct, and 'quasi-participatory accountability' denotes the mechanism through which we are able to recognise it.

Finally, I want to respond to some of criticisms of Kutz's approach that have been made by Iris Marion Young. Far from being decisive against the approach, I wish to suggest that Kutz's view could be used to enhance Young's own account of responsibility for collective harms, which she terms the Social Connection Model. Specifically, it can provide a richer justification for the dynamic of responsibility Young describes, in terms of a realistic moral psychology, which was something of a gap in her account. Young's critique of Kutz consists in the observation that it is not plausible that emitters intend to cause collective harms, and thus not appropriate that they should be morally accountable in a backward-looking sense for playing their part in producing them. As we have seen, however, while it is true that

individuals do not intend to cause climate change, it is a mischaracterisation of Kutz's view to claim that his account of quasi-participatory intention supposes individuals have this aim in particular.

She further worries that the attribution of guilt or blame in cases of collective harm is inappropriate because of the way such attributions tend to isolate individual perpetrators and imply the exclusion of others from accountability. A corollary of this is that attributions of blame or guilt express 'a spirit of resentment' and thereby produce 'defensiveness' (Young 2011, 114) when used in political discourse – people will tend to resist such attributions by pointing to others they consider worse offenders. Young suggests that 'what we should seek is not a variation on a weaker form of liability, but rather a different conception of responsibility altogether' (Young 2011 104). Young suggests we can and should adopt a revisionary conception of responsibility when considering collective harm. On this conception, participation in the structural processes that produce such harms should be thought of as generating a duty to engage in collective action aimed at dismantling those structural processes. This forward-looking responsibility (something approaching remedial duty, but less specific in its demands) can be viewed as a kind of outcome responsibility, in that it is grounded in some at least partly backward-looking considerations of involvement or participation. A reading of Young's view, which attempts to elucidate her account of the ground of responsibility, will be offered in the next chapter. What is clear is that Young's conception of social responsibility is not to be mediated by attributions of guilt.

Guilt, however, is also useful. It motivates one to make restitution, to reset the balance. It is not clear why mere involvement in collective harms should motivate one to engage in collective action to combat that harm, unless one regards oneself as morally "on the hook" for that harm. Guilt, from a first-personal perspective, and susceptibility to resentment, are

part of what it is to be on the hook. Young talks of her view as an alternative account of responsibility, one that we should adopt with respect to certain kinds of harm. The problem is, there are matters of fact about our practices of responsibility attribution that determine the conditions in which one can regard oneself as responsible, and no argument for the social usefulness of an alternative system will change those facts. The quasi-participatory intention approach can be seen as giving us an explanation as to why participation in harmful social-structural processes generates an obligation to campaign against those processes: a refusal to do so would indicate a continued participatory intention to endorse and preserve the conditions that give rise to those processes, and with it, complicitous accountability.

Participatory intention, it seems, cannot be dispensed with when considering the ways in which individuals can be judged to be "on the hook" for collective harms produced by groups of which they are members. It is for this reason that Cripps and Collins' accounts failed to provide convincing grounds for the attribution of responsibility or gap-filling duties. There is an obvious obstacle to the appeal to participatory intention to ground accountability for unstructured collective harm: that it is unrealistic since no such intentions exist. But, as we have seen, this objection can be answered: we can give plausible examples of participatory intentions of just the kind we need. A final worry might be that even if some emissions can be attributed to quasi-collective acts, a good deal of them cannot, since many carbon-intensive activities are literally unavoidable in a modern economy, even though people would like to avoid them if they could. With respect to such emissions, it may have to be conceded that there is no accountability gap to be filled: arguably, combatting climate change in relation to such emissions is simply a matter of collective benevolence.

5. Rethinking Moral Revisionism

So-called common-sense morality seems in some sense deficient when we consider the rights and wrongs of contributing to large-scale unstructured collective harms in general and climate change in particular. Dale Jamieson offers what has become one of the most famous expressions of this apparent deficiency: '[t]oday we face the possibility that the global environment may be destroyed, yet no one will be responsible'. He interprets this claim as indicating that 'our dominant value system is inadequate and inappropriate for guiding our thinking about global environmental problems', and that we must therefore 'develop new values and conceptions of responsibility' (Jamieson 1992). Climate Ethics is, as a discipline, unlike much of moral philosophy in this respect: 'climate ethicists are trying to get us to change our moral judgements rather than simply reporting them' (Jamieson 2014 7). The distinction between the ethicist and the moralist, between theory of ethics and its practice - taken for granted by the canonical philosophers of the Enlightenment - is jettisoned in this style of philosophy. Theorists like Jamieson are apparently asking climate ethicists to become moral revolutionaries, a role one might think better suited to preachers, prophets or poets. Is it right to dismiss the role of the more prosaic moral analyst?

Ethical revisionism comes in two forms, the partial and the radical. Partial revisionism consists in arguing that our existing moral concepts should be extended, that they should be applied in new contexts. In principle, it could also consist in the claim that our existing moral concepts have overreached themselves and should be withdrawn from certain domains, although this possibility might appear less appealing: it is most consoling to view

moral progress as a constant broadening of our horizons.[29] Revisionist arguments of this partial kind point to supposed phase shifts in the history of moral development – the inclusion of enslaved people into the category of bearers of human rights, the inclusion of women into the category bearers of political rights – and propose that we continue this inclusive expansion in new directions – for example by extending the judgment that certain harms against people are wrong to also include harm against animals, or against non-human nature as a whole. Of course, this picture of the history of ideas is tendentious – we need not accept that moral development has in fact been a straight march towards greater enlightenment. Nevertheless, the partial moral revisionist can still claim that such moral progress as there has been, wherever it occurs, has taken the form of a partial expansion of our moral concepts, even if we cannot point to a continuous march of progress throughout history.

Radical revisionism, meanwhile, is the idea that we should attempt to inculcate new values in ourselves, which have no obvious analogy in the contemporary ethical landscape, or which perhaps exist only in vestigial form as relics of ethical outlooks since abandoned. Radical revisionism encompasses those philosophers who believe that we should foster new virtues, valorising dispositions the general adoption of which will lead to a more harmonious relationship with our world. It also includes those philosophers who hold that we should recognise revisionary structures of accountability, structures of obligation that are now said to arise in situations where previously they would have been considered inappropriate. Radical revisionists are pioneers of new frontiers, but they do more than simply map the ethical landscape, they make it anew.

[29] For an argument that role of de-moralisation has in recent times been neglected, both in terms of its emancipatory power and its historical significance within the Enlightenment discourse of moral progress, see (Buchanan and Powell 2017)

Peter Singer is cited by both Jamieson and Judith Lichtenberg as one of the most ground-breaking exponents of the former tendency: in *Animal Liberation* he entreats us to see that our concern for the wellbeing of human beings should also ground concern for the wellbeing of animals, and in "Famine, Affluence and Morality" he calls on us to acknowledge that the reasons we have to rescue someone nearby, whose needs are affectively pressing, should still apply when those in need are spatially and emotionally distant from us (Singer 1975, 1977).

Lichtenberg and Jamieson themselves can be accounted as radicals: Lichtenberg describes the view that negative duties take priority over positive duties as the 'commonsense' view and presents reasons why this view is 'overrated' (Lichtenberg 2010 560-1), cajoling us towards the conclusion that we should abandon it. Jamieson argues that 'commonsense morality does not commit us to the views climate ethicists say we should hold', meaning that 'new moral understandings are required if we are to moralize some important aspects of our climate-changing behaviour' (Jamieson 2014 170). His key claim is that we should aim at 'nourishing and cultivating particular character traits' or 'green virtues' (Ibid. 186). Iris Marion Young's work on responsibility provides another example of the radical revisionist tendency: she argues that 'practices of assigning responsibility in...everyday moral life' are unsuitable when it comes to the question of responsibility for injustices which are emergent from the normal activities of many individuals acting together, termed 'structural injustice'. She therefore proposes we recognise 'a special kind of responsibility', which she names 'the social connection model of responsibility', as an alternative to 'the conception usually applied in legal and moral discourse' (Young 2011 95–7).

The partial and radical tendencies should not be viewed as mutually exclusive or even strictly delineated at all, as Jamieson recognises. Advocates of radical shifts in our moral outlook make their case from the perspective of our current practice, which means any

argument for radical revisionism will often begin with a call to extend or reframe our present attitudes. Partial revisions, for their part, may be viewed as a mere extension of our existing moral practice from the perspective of a particular class of moral theories, such as consequentialism, but may nevertheless appear radical from the perspective of another, such as virtue ethics. Yet despite these emollient considerations, radical revisionism of any kind is difficult to motivate, and may even seem to venture into territory that a certain tradition at least may not regard as the province of philosophy at all.

There is a tradition in philosophy which regards modern thought as being characterised by the view that 'moral principles and precepts are accessible to normal and reasonable persons generally', and that the problems of moral philosophy lie 'not [in] the content of morality but its basis' (Rawls 2000 10-11). This outlook is shared by most of the Enlightenment cannon, and is taken up by contemporary figures such as Rawls and his followers. For the philosophers of the Enlightenment tradition, the job of the ethicist was to map the foundations of common sense moral judgments, not to change them - or at least not to do so radically. Kant, for example, claimed that 'even the most hardened scoundrel' would recognise examples of right action and act accordingly, if only he could shake off immoral impulses, from which he 'wishes to be free' and are 'burdensome to him' (Kant 1997 59, AK 4:455).

Of course, common moral judgments can be wrong, and some may rightly be regarded as prejudices. The history of ideas shows us that many attitudes we now view as vicious biases were once regarded as correct moral judgements. Kant himself, for example, seemingly approved of slavery on the grounds that Africans were congenitally indolent unless forced to work (see Kleingeld 2007). It would be foolish to regard the morality of common sense as immutable or incorrigible. But it is still a mainstream view that moral philosophy should proceed as if common sense morally was at least moving towards the correct way of

thinking, and differing judgements about similar cases were to be explained by different interpretations of the facts on the ground. At the very least, we are often suspicious of philosophers who take ethics to be wholly creative enterprise, rather than an attempt to map the contours of a phenomenon that we encounter around us. Those philosophers, most notably Nietzsche and those he influenced, who saw their role as one of effecting evaluative shifts, had to work very hard to change philosophy in such a way as to make it adequate to this task. Without this work, as Bernard Williams noted, 'what is conceived of as a radical philosophy will unsurprisingly turn out to be just like conventional work which equally lacks intensity' (Williams 1989). The assertion that ethics ought to be radically different has a tendency to comes across as a failure to produce arguments that specific putative principles are false, or misapplied.

In recent years, a tradition of practice dependence has come to prominence in political theory, whose central contention is that political norms should not be justified on the basis of principles that are abstract and universal, but should rather be justified on the basis of principles that are determined by a particular institutional context. Practice dependence as a tradition in political theory can be viewed as having been presaged and pre-empted by a corresponding tradition in moral theory, exemplified by Peter Strawson's "Freedom and Resentment".[30] Christopher Kutz can be regarded as a contemporary inheritor of this tradition. 'Practices of accountability' Kutz writes, 'comprise a system for protecting and maintaining social interests', and his contention is that from this fact we can derive conclusions about the best interpretation of those practices. Specifically, the claim is that we should not conceive of accountability in a 'solipsistic manner', whereby praise and blame are treated as responses to merits and demerits logged in a metaphysical 'account'

[30] In drawing this parallel I follow (Jubb 2014)

(an image borrowed from (Feinberg 1970 124-5)), but rather should view accountability as a system of reactive attitudes that are appropriate from certain salient perspectives.

This chapter begins by giving an account of the Strawsonian approach to moral theorising, which views the system of reactive attitudes as a complex emotional dynamic that plays a foundational role in determining the appropriateness of our moral judgments. It defends that account against an alternative conception of the Strawsonian position, due to Christine Korsgaard. The following sections argue that the forms of radical moral revisionism defended by Jamieson, Lichtenberg and Young are incompatible with the Strawsonian conception of morality, and that this should give us strong reasons to be suspicious of them. But the pitfalls of revisionism should not be cause for despair, as, it will be argued, we already have the conceptual resources to morally engage with the problem of climate change in an adequate way, in a way that does not conflict with the participatory attitudes which constitute our existing moral practice.

The Strawsonian View

The image of moral worth as an account in which debits and credits are logged can be traced back much further than Feinberg: it occurs in certain translations of the Lord's Prayer - 'forgive us our debts, as we forgive our debtors' (Matthew 6:12, King James Version; cf. 'forgive us our sins, for we also forgive everyone that is indebted to us', Luke 11:4). On a certain reading at least, Kant can be viewed as an inheritor of this religious tradition, and Kantians are perhaps Strawson's target when he characterises the pessimistic character, who worries that determinism might undermine the possibility of just punishment, as when he characterises the pessimist as one who demands 'a genuinely free identification of the will with the act' (Strawson 2008 3) if one is to be justly punished for that act. This may be thought to correspond to Kant's claim that people could not be 'reproached as guilty of their crimes...if we did not suppose that whatever arises from a

man's choice has a free causality as its ground' (Kant 2015 80, AK 5:99-100).[31] It is this identification of the will with the act that is supposed to constitute a debit in our moral account. Practices of praise and blame, and correspondingly of reward and punishment, are an integral feature of the moral system, from this perspective. Blame and punishment, on this view, are hardly distinct from the idea of moral judgment: they are responses to moral desert.

In Sidgwick, a separation emerges between the system of morality and the practice of praising and blaming, or of partaking in moral dispositions generally. Part of Sidgwick's project in *The Methods of Ethics* is to identify respects in which utilitarianism appears to diverge from common sense morality, and to provide arguments minimising those areas of divergence, in order to give the most cogent version of a utilitarian theory. He finds one such gap with respect to the issue of demandingness. Normally, we blame people for failing to do what is right. Further, Sidgwick holds, if one judges that something is the best thing one could do in a given circumstance, and that one is able to do that thing, then if one fails to do it, one would have failed to do what is right. But we can think of many situations that fit this model where we do not blame the agent in question. For example, it may be best for 'a rich man' to 'live very plainly' and 'devote his income to works of public beneficence', but we do not typically blame rich men who fail to do this. To resolve this apparent conflict between the utilitarian imperative to do what is best and common sense morality, we are asked to consider 'the practical effects of praising and blaming'. The thought is that we contribute to moral progress and therefore the overall good more by restricting our praise to acts that are 'above the level of ordinary practice', and restricting our blame to those acts

[31] Kant, of course, only asserts that we are entitled to reproach people for their crimes because we are entitled to *regard* their choices as the product of a free causality. He does not claim that we have theoretical knowledge of the existence of such a causality; this would be a transcendental fallacy according to his system.

that fall 'clearly below that standard' (Sidgwick 2017 104).

This separation between the right and the praiseworthy, the wrong and the blameworthy, generates a certain tension: Sidgwick uses the principle of utility to determine who has failed to do what is right, and in that sense who is deserving of blame, but also appeals to the principle of utility to determine when blame should be withheld for someone who fails to do what is right. Blame is thus both a judgment that reflects the determinations of the principle of utility, and a practice that is justified by appeal to the principle of utility. Bernard Williams noted this tension, remarking that 'there is a deeply uneasy gap or dislocation in this type of theory between the spirit that is supposedly justified and the spirit of the theory that supposedly justifies it' (Williams 2006 289). What justifies moral practices like blaming, for Sidgwick, is quite distinct from what it appears such practices are for, from a perspective internal to the practice. Sidgwick's decision to preserve this tension is an aspect of his advocacy of esoteric morality, attacked as 'Government House Utilitarianism' by Williams (Ibid. 291). One problem with moral esotericism, in addition to the elitist attitude it involves, is that it might seem to undermine useful moral practices in the philosopher's own case. If we enlightened ones know that the real reason for blaming others is that doing so is instrumentally valuable, practices like blaming become a kind of charade, as when one feigns anger at a dog in order to train it. This conflicts with our experience of such practices, which have 'thickness': they are meaningful and are part of what give life meaning.

In "Freedom and Resentment", Peter Strawson makes the key claim that so-called reactive attitudes, including (but not limited to) praise, blame, resentment and gratitude are simply a natural feature of our living alongside others. Though we can suspend these attitudes in special cases in which we judge someone to be less than fully rational, we would not - and should not - suspend them even if it were shown conclusively that our actions were

predetermined, by God or causality. The implication is that our reactive attitudes constitute a discrete system, based in social relationships and independent of which 'metaphysical' theories turn out to be true. Strawson rejects the view of morality according to which we blame others because they *morally deserve* it, where moral desert is determined by a metaphysical feature of their will. He adopts instead a view of morality according to which practices of praise and blame take explanatory priority, and to say that someone morally deserves blame is just to say that our practices of assigning praise and blame dictate that blame is appropriate in the present case. In this sense, the practices of praise and blame are themselves the very matter out of which morality is built, rather than being a mere means of keeping score with respect to some more fundamental phenomenon: moral desert. Furthermore, because these practices are taken as foundational to morality, the question of what *justifies* them is misplaced, as is the utilitarian response that led to Sidgwick's tension.

What does it mean for praise and blame to constitute a system based in social relationships, which operates independently of 'metaphysical' claims about morality? Strawson elucidates this picture by drawing out the connection between the reactive attitudes of praise and blame with those of gratitude and resentment. When one is assaulted by another, this is injurious not only to one's body, but to one's conception of oneself as a being worthy of respect. Resentment is an expression of one's demand for respect, a response to psychological dissonance between the regard one has for oneself and the regard afforded to one by others. Resentment can thus be quelled only by some act of restitution, such as a public apology or the payment of compensation, which serves as an acknowledgement that the other's treatment of one was inappropriate given one's status as an equal, and restores one's standing both vis-à-vis the assailant and in the eyes of the wider moral community.

An assault from a child, or someone suffering from mental illness, does not stir resentment to the same degree, and does not seem to demand restitution. But we do not need to appeal to the idea that such people are less than fully *free* in order to explain this difference with respect to reactive attitudes (and indeed this would not be an effective explanation in any case). We can simply observe that being struck by a child who knows no better implies no slight against one's sense of self-worth; though one might be injured physically, one is not injured in terms of interpersonal standing, and therefore no recompense is required. One does not demand the same kind of respect from a child as one demands from a mature person, because the value of one's relationship with a child does not lie in mutual recognition of equal standing, but in recognising that a child is in need of education in order that she can one day be ready to become a fully-fledged member of the moral community. If one reproaches a child, it is not to express genuine resentment, but to teach her the importance of interpersonal respect. One takes an 'objective attitude' towards the child, in Strawson's terms, which is distinguished against the 'participatory attitude' one takes in one's interaction with someone regarded as a moral equal.

Blame or moral disapprobation is then just something like resentment from an impersonal point of view, or resentment for a slight not against one's own esteem, but against the standing of the system of principles governing behaviour which protects one's own esteem as well as that of others, and through which we all expect decent treatment from others. We do not blame, *pace* Sidgwick, in order to uphold a useful system of social rules; rather, our blaming is itself a feature of our investment in that system. Just as I naturally resent those who damage or undermine that which I value, I blame those who damage or undermine that which is of general value. From Strawson's account of the role of praise and blame, which begins as a response to the problem of free will and determinism, grows an approach to moral theorising in general: it is only by attending to the reactive attitudes,

he writes, that we can 'recover from the facts as we know them a sense of what we mean; i.e. of *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation and justice' (Strawson 2008 24).

To 'speak the language of morals', then, is to respond to the structures of moral sentiment as we actually experience them: they are to be treated as given, by our 'human nature and our membership of human communities' (Ibid. 17). This Strawsonian picture of morality as a given lived world of emotional responses preserves the Kantian concern that practices of blame and punishment be integral to the moral system, rather than being instrumentally justified, as on the utilitarian account. But it is incompatible with a Kantian approach with respect to the primary position Strawson gives to emotion in the explanation of our moral life. Christine Korsgaard's reading of Strawson gives an incomplete account of the irrelevance of determinism to questions of responsibility, precisely because, in trying to make the Strawsonian view more amenable to a Kantian framework, she attempts to render the view in a manner that dispenses with any appeal to natural sentiments (Korsgaard 1992). An examination of what is missing from Korsgaard's view will give us a clearer picture of the Strawsonian approach.

Korsgaard distinguished two stances one may adopt with respect to considerations of praiseworthiness and blameworthiness: the theoretical and the practical. These, it seems, roughly correspond to Strawson's objective and participatory attitudes. Like Strawson, she observes that it is sometimes appropriate to adopt one or other of these attitudes when considering questions of responsibility in specific cases, but we err when we attempt to apply one or other attitude to all cases. When we adopt the theoretical stance, we give explanations for phenomena in terms of the chain of cause and effect. When we adopt the practical stance, we give justifications for actions in terms of reasons. Strawson gave the label 'pessimist' to those who, mistakenly, apply the theoretical stance to the question of

whether people in general can be held responsible for their bad behaviour. The pessimist claims that one can always give an explanation for any piece of behaviour in terms of natural causes, and that this fact is always exculpatory (and is therefore pessimistic about whether practices of accountability make sense given the truth of determinism). To someone who adopts the theoretical stance globally, the only circumstance in which we would judge someone to deserve blame would be one in which there was some fact about her that "came between" the natural causes of her present state and her bad action, one that explained her bad action. This would have to be something like an unconditioned wicked will. When we apply the practical stance, however, facts about natural causes may be irrelevant when determining whether the behaviour in question was justified.

Korsgaard argues that excuses for bad behaviour function as 'practical reasons for not holding a person responsible' (Ibid. 313), rather than evidence that the agent was less than fully free, and on that basis agrees with Strawson that the thesis of determinism is in principle not the sort of thing that could undermine attributions of moral responsibility. Say someone behaved badly because she was nervous. Clearly the fact that she was nervous constitutes a reason to excuse her of her bad behaviour, in the sense that the fact she was nervous does indeed seem to be a consideration that favours excusing her bad behaviour. But here, Korsgaard's explanation stops. To understand nervousness to be a reason excusing bad behaviour, Korsgaard claims, 'we need only to know what it is like' (Ibid.). Nervousness is certainly the sort of thing that makes it harder to control one's emotions, and so might, for example, make one prone to violent outbursts. But some cases of failure to control one's emotions are blameworthy, while others are less so. When we try to explain why, we are immediately drawn to features of the case beyond just what nervousness is like: it seems natural to point out that nervousness is generally caused by a factor outside of oneself and in that sense beyond one's control. Conversely, it might be said, in the case of

an angry outburst where nervousness is not involved, one can only point to factors within one's own person, making one guilty. Obviously, if we allow ourselves to fall into this line of reasoning, we will have returned to playing the 'theoretical' game of looking for some feature of the individual that creates a debit in her moral account: her wicked will.

Kant and Korsgaard would apparently respond that this enquiry into the hierarchy between 'inclination' and the will is a dialectical mistake, because it demands, in Kant's terms, knowledge of the noumenal, which is impossible. Korsgaard's claim, in other words, is that explanations have to stop somewhere, and hers stops here. Nervousness is a reason to excuse lashing out at someone, while having an angry temperament is not, and the difference between the two cases has something to do with the different way in which the will interposes itself between inclination and action in each case, but more than this we cannot say. Strawson, however, has much more to say: he can give us a compelling description of the emotional dynamic that underlies the interaction between the nervous person and the recipient of their intemperate abuse. By protesting one's nervousness, one emphasises that one did not intend any slight against the esteem of the injured party, and so there is no cause for her to feel resentment. One points out that the feature of the case she resents, her implied abasement, is not in fact present. Indeed, the offering of the excuse itself plays a part in the emotional dynamic, similar to that of an apology. The act of offering an excuse is an act whereby one disowns the original bad behaviour, acknowledging that if one had intended any affront to the esteem of the injured party, one would indeed be worthy of resentment. One thereby takes common cause with the injured party, re-establishing that one shares the respect the injured party has for herself and for the norms of decent behaviour.

Korsgaard claims that the question of whether one adopts the theoretic or the practical stance, the objective or the participatory attitude, must be more than a matter of 'mere'

emotion, as sometimes one cannot help reacting angrily to pesky children and broken coffee machines, even though one knows that the objective stance is the appropriate one in such cases. In these cases, she thinks, one chooses to adopt the theoretical stance in spite of one's reactive attitudes, or as she puts it 'the feelings that accompany' one's reactive attitudes (Korsgaard 1992 321). But Strawson is not arguing that just any emotions can constitute the relationship of holding someone responsible. Anger and frustration alone will not do the job. Rather, he is describing a particular interplay between the attitudes of self-love, offense, and resentment. If one truly resents a coffee machine when it eats one's change, one treats it as the sort of thing whose esteem is of value. Given a coffee machine is incapable of esteem one would be straightforwardly confused. The same is true, in a qualified way, of resentment for a child. If one feels genuine resentment against a child for some perceived slight, it follows one values her esteem. It is possible that the particular nature of one's relationship with the child means this is not a mistake: the extent to which it is appropriate for one to resent a child is a matter of degree, proportional to her degree of social standing and maturity.

Imagine, for example, a child monarch, whose favour is of great social importance. A refusal to grant such favour will understandably be met with genuine resentment. This is not, *pace* Korsgaard, best explained by the idea that one chooses to treat the child monarch as a full member of the moral community – why this child and not others? Rather, it is best explained by the fact it is appropriate for one to value the child's esteem. Strawson can also give a much better account of tricky cases of disagreement about blameworthiness. Take the case of a criminal who developed a tendency towards violent behaviour because of a history of being the victim of abuse. Those people who are inclined to the view that the criminal's history is not exculpatory can be inferred to place a high value on the criminal's showing the proper degree of esteem for other members of the moral

community, for social norms of non-violence. Those who are inclined to excuse the criminal, conversely, can be viewed as judging that, because the criminal has been in some sense morally damaged by historic abuse, it is of less importance whether the criminal shows the proper degree of respect towards his fellow citizens. They value the esteem of a damaged individual less, because it is less meaningful, just as the esteem of children is less meaningful. The criminal is in effect judged incapable of participating in the kind of relationship with the moral community that they consider valuable, meaning the criminal behaviour cannot undermine that relationship. This does not mean they treat the criminal as being of lesser moral worth; indeed, once they adopt the theoretical/objective point of view, their arguments for clemency are likely to invoke the criminal's interests as a moral person, such as the possibility of reformation.

The Strawsonian picture of morality, then, consists in a system of reactive attitudes, conceived as natural emotional responses, which appear to us as the given structure of social relations. What we mean when we talk about moral responsibility and justice is 'recover[ed] from the facts as we know them', and the relevant facts in question are the facts about our reactive attitudes in given cases. In what follows, it will be argued that this conception of morality sits uncomfortably with the idea that climate change, or anything else, makes it the case that we should be radical moral revisionists.

Revisionism about Virtues

Jamieson's radical revisionism consists, to reiterate, in the claim that we should aim at fostering novel green virtues. Virtues are for Jamieson a device for bringing the results of the best moral theory into harmony with plausible moral psychology. Utilitarianism, Jamieson believes, gets something fundamentally right at the level of moral theory, in that it requires us to do what is best. But it gets something fundamentally wrong as a formula for living a moral life. The mistake lies in assuming that, having accepted the utilitarian

principle, the next question should be what kind of utilitarianism is the right one. Varieties of utilitarianism – act, rule etc. – specify a 'level' (Jamieson 2007 170) of evaluation, a particular phenomenon that is to be the object of evaluation in terms of the consequences to which it gives rise. The usual form of criticism against these level-specific or 'local' (Ibid.) views is to accuse them of undermining the spirit of the utilitarian principle itself. If we choose acts as the object of our evaluation (if act utilitarianism is used a criterion of rightness), we find acts that are judged right even though they lead to bad consequences in the long run when iterated many times. If we take rules as the object of evaluation, meanwhile, we are left with rules that, while producing the best consequences of any possible set of rules overall, seem obviously to lead to bad consequences in exceptional cases.[32] Jamieson, at least in (Jamieson 2009) endorses global consequentialism as the best way of overcoming these level-specific tensions (Jamieson 2009 170, citing Parfit 1984 25, see also Driver 2001). We are to think in terms of a utilitarianism not of acts or rules, but literally everything: all things, from dispositions and intentions to eye colour, should be organised so as to produce the best consequences.

It is not immediately clear whether Jamieson means to reject 'local' utilitarianisms as criteria of rightness, or only as decision procedures, and which of these functions his preferred global utilitarianism is supposed to serve. It is not obvious how global utilitarianism overcomes the problems associated with both act and rule consequentialism when all these views are treated as criteria of rightness. Under global utilitarianism, acts and rules are *both* to be considered the proper objects of evaluation, along with everything else. Some act could be optimal as an act, but be ruled out by the optimal rules, the product of suboptimal motives or a manifestation of a suboptimal disposition, etc. This would seem to multiply

[32] For an argument that this form of criticism is not fatal to Rule Consequentialism, see (Hooker 2000 Ch. 4).

the potential for conflict under the utilitarian principle broadly conceived, rather than resolving conflicts. An important part of Jamieson's opposition to level-specific utilitarianism is his opposition to 'calculation', founded on the widely accepted premise that it is near impossible for individuals to calculate the badness of the consequences of (say) their actions, given the limited information and computational abilities most people have available to them (see e.g. Lenman 2000). But again, it is not clear how global consequentialism resolves this concern: the global consequentialist multiplies the number of features of situations whose consequences have to be 'calculated'.

It seems most coherent to read Jamieson's opposition to calculation as consisting in opposition to any level-specific form of utilitarianism being treated as a decision procedure, while endorsing them all, in combination, as criteria of rightness. Is he then endorsing global utilitarianism as a decision procedure? Global utilitarianism treated as a decision procedure would seem to be impossible, as various dimensions of evaluation would very likely conflict with one another. But Jamieson could respond that this was precisely the point. The injunction to treat global utilitarianism as a decision procedure would be tantamount to an injunction to abandon the idea of a decision procedure altogether. [33] In the place of a decision procedure, we are to look to 'noncalculative generators of action' (Jamieson 2007 167), and it is in this capacity that the virtues enter the picture. The motivation to act in accordance with virtue, Jamieson claims, has the feature of 'non-contingency', meaning it is not dependent on predictions about the behaviour of other

[33] Driver (2011 176) gives a similar response to Hooker's criticism of global consequentialism (Hooker 2016, §5). Hooker points out that a global consequentialist would have to be a consequentialist about both decision procedures and acts. Say consequentialism applied to decision procedures showed procedure P was best, and that P recommends act A. Suppose further that consequentialism applied to acts showed not-A was best. Global consequentialism would apparently recommend both A and not-A. Driver responds that this is not a paradox, but a reflection of moral ambivalence: A is choiceworthy in one way but not in another. Jamieson can similarly be read as rejecting choice algorithms.

agents. The motivation to perform some act because it is recommended by an act consequentialist decision procedure, meanwhile, instantiates 'contingency', it is dependent on the behaviour of others. Predicting and evaluating the possible consequences of the behaviour of others requires calculation, and calculation is difficult. This is why, Jamieson thinks, 'utilitarians should take virtues seriously' (Ibid).

What exactly is it, then, to 'take virtues seriously'? Jamieson does not consider virtues to be of particularly deep significance, being effectively subordinate to the principle of utility, on his view. Presumably, this explains why he seems equally untroubled by the question of exactly which traits should count as virtuous, and how this list of traits is to be compiled. He sees virtues as falling into the three categories, those of preservation, rehabilitation, and creation (Jamieson 2014 186). These correspond respectively to virtues drawn from existing values, virtues that combine 'additional or different content' with existing values, and virtues that reflect new values. An example of the first is humility, including the humility to 'not destroy redwood forests'. An example of the second is temperance, a traditional Christian virtue that Jamieson claims can be rehabilitated to incorporate the idea of reducing one's consumption of environmental resources. An example of the third is 'mindfulness', which Jamieson claims would involve being conscious of the ways in which one's actions have remote consequences in time and space, and thus 'taking on the moral weight' of the consequences of production and disposal of items we buy.

This tripartite division seems to indicate that virtues are to be drawn both from tradition and from some kind of direct consequentialist evaluation, as well as a combination of the two. The rationales for the merits of these sources push in opposite directions. Presumably, if we should draw virtues from tradition, it is because doing so will produce the best consequences. Jamieson does not argue this expressly, but it is easy to see how such an argument would go: building on moral dispositions and attitudes that are already well

established is likely to be much easier than trying to change peoples' attitudes entirely, so we can expect preservation and rehabilitation to benefit from more widespread uptake as methods of behavioural change. If we are to take virtues from a utilitarian evaluation of various possible character traits, meanwhile, it must be because *this* will produce the best consequences. The two strategies are not straightforwardly contradictory; it could be that preserving traditional virtues produces the best consequences in certain cases, and discovering new virtues through the evaluation of traits does so in others. But distinguishing when we are in one situation and not the other would seem to involve very difficult calculation. If avoiding the need to carry out such calculations was supposed to be the main reason consequentialists should 'take virtues seriously', it looks like Jamieson's consequentialism undermines his virtue ethics, and *vice versa*. Jamieson's response here would apparently be that 'there is no algorithm for designing the optimal utilitarian agent', but that 'we sometimes know them when we see them' (Jamieson 2007 180). But this appeal to common sense is totally at odds with his revisionist aspirations. If we all already know virtue when we see it, environmental virtue included, the initial claim that our prevailing moral framework is fundamentally unsuited to the contemporary world and in need of radical reformation cannot also be true.

More importantly for our purposes, Jamieson's view runs up against the same kinds of considerations Strawson and Williams raised against approaches like Sidgwick's. In *Reason in a Dark Time*, Jamieson offers an extended analysis of the putative virtue of respect for nature. He gives three reasons for respecting nature: that it serves self-interest to apply precaution in our interaction with nature, that nature provides the background conditions of a meaningful life, and that respect for nature 'flows from...psychological integrity and wholeness' (Jamieson 2014 191). Taken in the context of Jamieson's wider theoretical framework, this list of reasons implies a 'deeply uneasy dislocation' between 'the spirit that

is supposedly justified and the spirit of the theory that supposedly justifies it', to repurpose Williams's criticism of Sidgwick. Clearly there are good reasons to respect nature, which are part of what makes respect for nature a practice that is deeply meaningful and important in the lives of its practitioners. But these reasons (with perhaps the exception of the appeal to precautionary reasoning) sit very uncomfortably with the idea that the ultimate justification for the practice is to be found in the principle of utility. Jamieson is clearly quite correct to point out that the natural world plays an importantly meaningful role in the lives of the communities that grow up within it – he mentions, by way of example, William Blake's vision of 'England's green and pleasant land', which was meaningfully intertwined with Blake's spirituality and national identity. But this case actually helps to bring out the problematic dislocation. Blake's intrinsically valuable relationship with geography has precisely nothing to do with utility. Rather, it is internally justified; it is self-justifying. The relationship would continue to structure Blake's motivation whether or not utilitarian considerations ultimately supported its doing so.[34]

Jamieson's approach requires him to bridge the deep gulf that lies between two sources of moral motivation. These sources correspond to Strawson's objective and participatory attitudes. Let us continue to consider the psychology of the figure of William Blake. Blake takes a participatory attitude towards nature, or more precisely, nature features in participatory relationships in which he partakes. His esteem for those who respect nature - which must be regarded as part of what constitutes respect for nature as a virtue - is a feature of the emotional dynamic that structures those valuable relationships in which he is engaged. Perhaps, for example, for Blake showing esteem for those who respect nature is part of how he engages in the practice of religious worship and signifies fraternal affiliation.

[34] The figure of William Blake is here used as a stock character for philosophical example, no claims to historical accuracy about the biography of William Blake are necessarily implied.

Viewing respect for nature in utilitarian terms, however, does not arise out of participatory relationships with the people around one in any immediate way. If one did not consider respect for nature a virtue in the way that Blake did, it is very difficult to see how Jamieson's work would serve to induct one into that practice. To those who do not already participate in the practice of treating respect for nature as a virtue, the reasons Jamieson cites for doing so function more like explanations – they give us insight into why someone would participate, but they will not necessarily be motivating in themselves.

Bridging the gulf is thus the difficult task Jamieson faces. He can give a good argument as to why it would be *better* if people recognised and were responsive to green virtues, but it is not at all easy for him to construct an argument that would *rationally persuade* people to recognise and be responsive to green virtues, if they did not already do so. Nor has he given a plausible description of how a shift towards an ethic of green virtues could be effected through non-rational means. Commenting on the nature of moral progress, Jamieson remarks that 'even radical critiques of existing practices are to a great extent immanent', and that 'this progress is often surprisingly tentative, incremental, localised, and even one-dimensional' (Jamieson 2016 12). If this is right, Jamieson arguably has to admit that a shift towards an ethic of green virtues can only be expected to happen as the end-point of a whole series of piecemeal adjustments to our present moral outlook. Read in this light, Jamieson's advocacy of green virtues is not so much an attempt at rational persuasion as an attempt to expand our imaginative horizons: it is a vision of a positive future state. But the work of getting there has to be done through reasoning internal to our existing moral practices, through adopting the participatory attitude. Jamieson, we have to conclude, has not much advanced us in this project.

Revisionism about Positive and Negative Duties

The foregoing discussion of Jamieson's view has given us strong reasons to be suspicious

of the idea that we can use abstract theoretical concerns to determine our reactive attitudes in our everyday moral relations with others. The recognition that holding certain traits to be virtuous would be socially beneficial is just not the sort of consideration that has a bearing on whether we actually hold people who manifest those traits in esteem. This result supports the claim we are pursuing, namely that the same kinds of concern that made Strawson sceptical of the optimist should make us sceptical of the moral revisionist. Further evidence for this claim can be drawn from an analysis of Judith Lichtenberg's proposal for another form of moral revisionism intended to respond to climate change, an example of what she calls 'New Harms' (Lichtenberg 2010; 2014 Ch.4).

Lichtenberg's project in "Negative Duties, Positive Duties and the "New Harms"" is to critique the traditional view that negative duties to avoid doing harm are more stringent than positive duties to help mitigate harmful outcomes, and to show that an appreciation of so-called New Harms should lead us to abandon it. New Harms are defined as harms caused by the contribution of many people, where no one individual can be said to cause the harm.[35] Her argument for revisionism seems deceptively moderate. She argues first that although it is clear that someone who has caused suffering has an additional reason to help alleviate it, in comparison to a bystander, the bystander may also have strong reason to alleviate it. She observes that in the case of New Harms, positive actions (such as donating money) may alleviate suffering more effectively than refraining from contributing to harm. Thus, for example, it might be better to donate to charities which aim to help people escape forced labour mining precious metals in central Africa, than to try to boycott electronics that contain such metals. Lichtenberg argues that considerations of integrity support positive duties just as much as negative duties. She further argues that objections of

[35] The term 'collective harm', the preferred term used in this thesis, is apparently a basically equivalent concept.

demandingness against positive duties can also be said to apply to negative duties, meaning the asymmetry between the two is overstated. Despite the apparently modesty of her claims, her view has been read as a call to arms for the radical moral revisionist, and has helped to initiate a literature that has recently been dubbed the 'New Harms Discourse' (Peeters, Bell & Swaffield 2019).

We should distinguish the elements of her view that are not radical from those that are. Part of her argument is simply that while in general, one has stronger reasons to avoid causing a certain amount of harm than one has to alleviate an exactly equivalent amount of harm, because one's individual behaviour may make no difference with respect to New Harms, and because one might make a significant difference through positive action, one may have stronger reasons to take positive action than one has to avoid the kinds of behaviours that, in aggregate, produce New Harms. But this should come as no surprise – if my behaviour does no harm, then the negative duty to refrain from doing harm gives me no reason to avoid it. However weak my reasons are to take positive actions, they must clearly be stronger than having no reason at all.

The radical reading of her view is that in response to the increasing prevalence of New Harms in the modern world, we should recognise a duty to aid that is as stringent as our duty not to cause determinate harm. At a first pass, the view would be that, with respect to New Harms, one should prioritise one's actions aimed at alleviating harm according to the amount of difference one can make, rather than the degree to which one is causally responsible for the harm in question. '[I]t seems likely', Lichtenberg writes, 'that, per unit of human effort (measured in dollars, or some other way), we are more likely to make a difference by giving aid than we are by refraining from contributing to harm', the suggestion being that this counts in favour of prioritising aid over refraining from contributing. This in turn suggests that, even if one does individually make a difference to harm through

contribution to New Harms, it is more important to focus on doing good that it is to avoid doing harm. The idea that effort is being treated as a scarce resource must also be important – why should we prioritise aiding over not harming when we can do both? The best formulation of the radical thesis, then, is as follows: if, for a given cost (a given amount of effort), one can do more good by taking positive steps to alleviate a New Harm, than one can by refraining from contributing to the New Harm, one should help to alleviate suffering before one refrains from contributing; and if one cannot do both, one should prioritise helping.[36]

This is indeed a revisionary claim from the perspective of common-sense morality. As Lichtenberg notes, 'since with the New Harms an individual's actions do not produce palpable, immediate, visible effects, one is likely to feel no regret, no guilt, no shame, and no drive to act differently' (Lichtenberg 2010 561). Lichtenberg acknowledges, then, that recognition of the New Harms is not at present supported by our practices of accountability; it is not expressed through our system of reactive attitudes. As the discussion of Jamieson has already suggested, we might think that this fact in itself presents a formidable obstacle, which it is difficult to see how mere argumentative persuasion could circumvent.

Consider an argument Lichtenberg makes on the basis of the value of integrity. The argument is illustrated by means of Williams's case of George the chemist (Smart and Williams 1975 97). George reasons that it is permissible for him to accept a job manufacturing chemical weapons, because if he declined the job, someone else would accept it, meaning his refusal to accept the job would make no difference in terms of harm.

[36] It is not clear from Lichtenberg's work whether, in her view, if one prioritises helping to alleviate suffering over contributing to harm, it must be the same harm in both cases, and indeed what the criteria of identity for harms would be if so. It is not necessary to settle this question for the purposes of the present critique.

Williams argues George is wrong to conclude that he may take the job, as 'each of us is responsible for what *he* does, rather than what other people do'. This case seems inimical to Lichtenberg's aims, indeed, in the same essay Williams remarks that we should draw a clear moral distinction between 'those things that I allow or fail to prevent' and 'those things that I myself, in the more everyday restricted sense, bring about' (Smart and Williams 1975 95). But, Lichtenberg claims, a concern for integrity, 'the expressive function of one's conduct' (Lichtenberg 2010 570), should be viewed as applying equally to omissions as it does to acts. On this view, one fails to show integrity if one fails to do one's fair share of harm avoidance, defined as one's share of whatever actions would 'appropriately relieve need' (Ibid. 571).

This appeal to fairness changes the subject. A duty of fairness is quite different from a duty of aid, in that a duty of fairness is owed to the other members of some standing system of social cooperation. A failure to give someone what is owed to them can be viewed as a negative duty – giving what is owed is a moral baseline, and one has a negative duty to refrain from deviating from that moral baseline. Put another way, a negative duty is a duty to refrain from wronging others, and failing to do one's fair share would arguably wrong the other members of the scheme of cooperation in which one was involved. Either such a scheme of cooperation exists, or it does not. If it does exist, then our reasons for taking steps to alleviate harms do not arise from a priority of positive over negative duties, but from the priority of one kind of negative duty over another. If it does not exist, Lichentenberg has not given us sufficient reason to suppose that individuals have stringent duties to help alleviate harms.

There is thus a serious explanatory gap in Lichtenberg's argument. She gives us good reasons to think that the case for negative duties with respect to New Harms might be weak (no-difference considerations), and she gives us good reasons to think that our positive

duties to aid might be comparatively stronger. But what she lacks is a strong case for the claim that the New Harms give rise to positive duties that are very stringent in absolute, rather than comparative terms, duties of a degree of stringency similar to uncontroversial cases of negative duty – not to kill, steal etc. The revisionist aspirations implicit in the New Harms concept are not matched by correspondingly revolutionary arguments. The hidden steps in her reasoning seem to be that because New Harms seem to involve the sort of effects that are usually prohibited by negative duty - exploitation, the imposition of avoidable suffering, avoidable death - and because negative duties have proven themselves no longer up to the job, positive duties should come in to fill the gap. Like Jamieson and Sidgwick, then, Lichtenberg is attempting to invoke higher-order theoretical considerations to support the claim that we should modify our system of reactive attitudes. But this claim involves the objective stance – it is essentially an argument for a kind of non-rational training, the modification of our affective responses in pursuit of the greater good. Lichtenberg is therefore fighting a losing battle, as no amount of rational persuasion will produce this effect.

Revisionism about Moral Responsibility

If we are unable to modify our attitude to positive and negative duties in the face of climate change, perhaps we can change the way in which we understand responsibility? Like Jamieson and Lichenberg, Iris Marion Young believes that problems like climate change present a challenge to our everyday moral concepts. Like Lichtenberg, and unlike Jamieson, Young does not make a great show of her radicalism or revisionary aims, indeed, she makes reference to a historical body of work on the concept of political responsibility, notably in the work of Hannah Arendt, the better to frame her own view within the context of existing moral practice. Nevertheless, there is indeed a radical revisionary core to the case she is trying to make.

Young's work on responsibility grew out of a series of papers on global labour justice and the sweatshop industry (Young 2001, 2004, 2006b), and culminated in the posthumously published *Responsibility for Justice* (2011). Like Jamieson's, her motivation for revisionism may be expressed through the thought that there might be certain forms of moral wronging – specifically injustice - for which no one is apparently responsible, and that this should strike us as problematic. Thus, for example, the negative impacts of Hurricane Katrina, which hit coast of the United States along the Gulf of Mexico in 2005, fell disproportionately on poorer African American communities, and this was not a matter of brute bad luck, but the result of many varied ways in which these groups had been marginalised as a result of people's avoidable actions. We should therefore characterise their plight as injustice. Yet many or most of the actions that constitute this unjust social structure are not blameworthy. Rather, Young wants to say, they are the shared responsibility of participants in those structures (Young 2006a).

As with Lichtenberg, there is a more and a less radical reading of the claim Young wants to make. The idea that it is the responsibility of citizens collectively to combat unjust social structures can be thought of as arising from a well-established tradition in liberal political thought, given the widely accepted premises that it is the responsibility of citizens to support the state, and that it is the responsibility of the state to secure justice, at least to some minimally acceptable standard. As Young sets out her mature view, however, she is making a different claim, namely that it is the responsibility of all those who are involved in unjust social structures to work towards ending them, whether or not this work can be mediated through established institutions such as states. Importantly, Young believes involvement in unjust social structures grounds forward-looking responsibility on the part of individuals, even if those individuals cannot reasonably be assigned backward-looking accountability, or liability, for those injustices. What is left unclear by Lichtenberg is made somewhat more

precise by Young: there is indeed some relationship between the "failure" of negative duties to prohibit certain group-caused harms and injustices, and the claim that positive duties should be thought of as stringently applying. The positive responsibilities we take on are duties of *justice*.[37] Duties of justice are standardly regarded as more stringent than duties of beneficence. These duties arise, on Young's account, from what individuals *do*: they arise from individuals' "social connectedness" to structural injustice.

Young's view stands in need of further precision. What exactly does "social connectedness" to structural injustice consist in, and why should this relationship give rise to positive responsibilities? A certain amount of reconstructive work is required to determine how Young would most cogently respond to this question. First, structural injustice is itself defined as a situation in which individuals' options are unfairly constrained, or they face significant deprivation, and this situation arises neither from wrongs perpetrated by specific agents, nor from specific pieces of bad policy, but from 'social structural processes'. Meanwhile, others derive significant benefits from those same processes. 'Social structural processes', in turn, are understood as 'objective social facts experienced by individuals as constraining and enabling' and 'macro-social spaces in which positions are related to one another', which exist 'only in actions', and commonly involve 'the unattended consequences of the combination of actions of many people' (Young 2011 53). This description is not supposed to be a definition of social structure; Young implies such a definition may be impossible. We can perhaps instead view this as something like a characterisation of an ideal type.

[37] Young prefers to say individuals take on "forward-looking responsibility" with respect to structural injustice, rather than "positive duties" because, following Feinberg 1970, duties are understood as requirements to perform specified actions, whereas Young wants responsibility to describe a broader kind of obligation, one that can be discharged in multiple ways according to the judgment of the agent in the circumstances.

Young offers potentially conflicting indications as to how the idea of 'social connection' to structural injustice should be defined. She writes, '[i]ndividuals bear responsibility for structural injustice because they contribute by their action to processes that produce unjust outcomes' (Young 2011 105). This language of 'contribution' seems to ground social connection in a causal relationship between individual behaviour and injustice. In the very next sentence, however, Young claims that '[o]ur responsibility derives from belonging together with others in a system of interdependent processes of cooperation and competition through which we seek benefits and aim to realise projects' (Ibid.). Here, conversely, her account of responsibility for justice looks like a version of the cosmopolitan view of global justice, on the model of early Charles Beitz, whereby justice is a property of the 'basic structure' – the norms that regulate systems of social cooperation – and systems of social cooperation are understood as transnational commercial networks (see e.g. Beitz 1979). If this reading is right, individual responsibility for structural injustice would ultimately be grounded in considerations of fairness. It is perhaps wrong to accept the benefits of a scheme of social cooperation if one does not also accept responsibility for the burdens that scheme imposes on others – at least, this would seem to be the kind of claim to which Young is appealing.

Young also cites Arendt's conception of political responsibility as an important influence on the conception of responsibility she wishes to defend. Arendt wanted to resist the idea that Germans could be considered collectively guilty for the crimes of the Nazi regime, arguing that 'where all are guilty, none are'. An attribution of collective guilt would – wrongly – seem to diminish the individual blameworthiness of Nazis who had perpetrated specific crimes. Rather, Germans were responsible for failing to do more to arrest the Nazi political project. On Young's reading of Arendt, this form of responsibility does not arise from some vicarious accountability for the actions of one's co-nationals, but rather from 'a

duty for individuals to take public stands about actions and events that affect broad masses of people' (Young 2011 76). It arises from one's duty not to shrink from one's status as a politically self-determining agent.

These readings conflict not only with each other, but with other claims Young apparently regards as important. Young's conception of the Arendtian demand for radical political engagement is forward-looking rather than backward-looking; all people are subject to this demand, simply in virtue of being 'aware moral agents who ought not to be indifferent to the fate of others' (Young 2011 92). Young, however, seems to want "social connection" to be at least in some sense backward-looking: she seems to want it to be constituted by what one does, or by some contingent role that one holds. This might be one's 'contribution', as the causal reading brings out, or by one's being unfairly advantaged by unjust structures, as the cosmopolitan reading brings out. The causal reading cannot possibly be the whole story, as if an individual could be said to be the cause of injustice, that individual would straightforwardly be blameworthy, and the social connection model would not come into play. The cosmopolitan reading, meanwhile, seems to neglect Young's clear concern for individual agency: it is clear that she takes social connection to generate forward-looking responsibly for individual participants in unjust social structures, not just suitably placed institutions such as states or international governmental and non-governmental organisations.

A coherent reconstruction of Young's concept of social connection has recently been offered by Maeve McKeown. Individuals are socially connected to structural injustice when they 'reproduce the background condition in which they act' (McKeown 2018). In other words, individuals are socially connected to structural injustice, and thereby have forward-looking responsibility for combatting that injustice, if their behaviour constrains the behaviour of others in ways that, combined with the behaviour of many other agents,

produces structural injustice. The distinction between 'reproducing' the conditions of injustice, and causing injustice, can be cashed out as follows: following Anthony Giddens, Young takes the reproduction of social structures to occur when one instantiates the same 'positional relations of rules and resources' (Young 2011 60) that one presupposes when trying to realise one's goals. Reproducing injustice, then, is playing a certain role, a role that, while not necessarily causing anyone to be *unjustly* constrained in their options, does condition the behaviour of others to a certain extent. Injustice arises when this role interacts with the roles of many other agents. When the actor in question takes the whole system of other agents occupying particular roles into consideration in her decision-making, the degree of constraint on her choices does rise to the level of injustice. Giddens uses the example of speaking an English sentence as an instance of *reproduction* of the English language (Giddens 1979 77, cited in McKeown 2018). By speaking an English sentence, one does not cause the English language to exist – it already exists. But when one occupies the role of a speaker of English, one conditions the way others speak English: one provides an example of the norms that others are to follow. And the English language is constituted, ultimately, by nothing more than the way in which speakers of English speak. Similarly, a participant in the garment industry might be constrained by their role into performing actions that help to constitute the marginalisation of workers in that industry. A purchasing manager for a clothing chain would, for example, face an institutional pressure to get the cheapest possible price; this pressure in turn partly determines the poor working conditions faced by production-line workers.

Even this quite plausible account of how the relation of social connection to structural injustice could be non-causal leaves much unexplained, however. It does little to clarify *why* social connection should give rise to remedial responsibility. Young's revisionism about responsibility arguably faces an explanatory gap similar to the one found in Lichtenberg's

account of revisionism about positive and negative duties. Either Young's case for revisionism is too modest to match her professed aims, or she is advocating a kind of revisionism by fiat, in which case her account is vulnerable to the now familiar Strawsonian challenge.

Why should someone who reproduces structural injustice have responsibilities with respect to that injustice, responsibilities that someone who had not played a part in reproducing structural injustice would lack? In response to the question of how the burdens of shared responsibility for structural injustice are to be distributed, Young offers four 'parameters of reasoning' – power, privilege, interest and collective ability (Young 2011 142). The question of burden sharing is a separate issue from the one we are considering – Young first needs to convince us that individuals have responsibility with respect to structural injustice before addressing the question of what should count as adequately discharging that responsibility. However, some of the remarks she makes under these sections are suggestive of a justificatory case for individual responsibility, as are certain remarks she makes elsewhere. In what follows, a reconstruction is given of candidate justifications for individual responsibility for structural injustice. As will be seen, none is fully up to the job.

Power/collective ability: it is plausible that individuals that have the power to effectively combat structural injustice, and members of groups that have a high level of collective ability to do so, have more responsibility for combatting structural injustice than individuals who do not*, ceteris paribus*. This fact alone, however, cannot justify the link between social connection and special responsibility, as many individuals who are socially connected to structural injustice lack power and collective ability, and many of those with a high degree of power or collective ability to fight injustice may lack social connection. It is perhaps difficult, in the modern world, to find an example of a person who is both very wealthy and very isolated from unjust global markets, but this is a contingent matter. The fact that we

can imagine an Atlantean King, whose vast wealth was drawn from an economy that was entirely disjoint from the various global patterns of trade that instantiate structural injustice, is enough to demonstrate that those with a high degree of power to combat structural injustice may lack significant social connection to it.

Fairness/privilege/benefit: as already intimated, one might think that it is unfair for someone to benefit from their role in a social structure without also accepting responsibility for those people whose role in that same structure places them at a disadvantage. One problem with this explanation is that it would not explain how people who did not benefit from social structures incurred remedial responsibilities. This would tend to preclude responsibility for core cases of structural injustice. Take Young's central example of the garment industry: consumers of clothing whose producers are subject to injustice arguably do not benefit from those industries, insofar as they are not necessarily better off than they would be without those industries. Consumers need to clothe themselves, and so purchase whatever clothing is available. They may have no self-interested preference as to whether the garment industry is globalised and unjust, or locally based and fair.

Young invokes a principle that 'persons and institutions that are relatively privileged within structural processes have greater responsibilities than others to take actions to undermine injustice' (Young 2011 145). She does not defend the principle explicitly. This principle is about burden sharing rather than establishing responsibility, but if it holds, a further principle looks to be in same intuitive ballpark, namely that one is responsible for structural injustice *if and because* one is relatively privileged by it. Although these principles have a certain *prima facie* plausibility, they could be denied. For example, some would argue on voluntaristic grounds that it is wrong to assign special responsibilities to individuals who

have no opportunity to avoid taking on such responsibilities.[38] More substantive argument is therefore required if these principles are to be accepted.

Young's privilege principle might be regarded as similar to the "beneficiary pays principle" (BPP) as invoked in discussions of burden sharing with respect to climate change, insofar as both principles assign remedial duties on the basis of being relatively advantaged by a particular industry, rather than considerations of causal connection to harm. Much important work regarding the BPP has been done subsequent to Young's death, which may explain why she did not make this connection explicitly. It is therefore plausible that arguments for the BPP are a sensible place to look in order to reconstruct a privilege-based justification for the claim that social connection generates remedial responsibility. A variety of justifications have been offered for BPP, for example that one fails in a duty to condemn injustice when one accepts the benefits of injustice (see e.g. Butt 2007), that title to benefits is invalid when the transfer involves rights violations even when they are committed by parties other than the recipient (see e.g. Goodin and Pasternak 2016), or that retaining benefits may contribute to the persistence of wrongful harm (Barry and Wiens 2016). Many of these justifications would not overcome the voluntaristic objection, as what they rule out, to be precise, is the *acceptance* or *retention* of benefits – a voluntary act. It is not obvious that they entail remedial responsibilities when benefits are received unavoidably and cannot realistically be disgorged (for example because one depends upon clothing, or upon fossil fuels, for one's basic needs).

Doubtless, more could be said in defence of the application of BPP-considerations to the question of responsibility for structural injustice. For Young's purposes, though, the fact remains that many of the people she wishes to claim share remedial responsibility for

[38] For discussion of voluntaristic objections to special responsibilities being grounded in relationships or roles, see Sheffler 1997.

structural injustice are not the beneficiaries of that injustice, or are not privileged by injustice. If it is correct that 'reproducing' the background conditions of injustice is the best reading of social connection, then it must be noted that there are core cases of such reproduction in which the agent is not privileged. The central example would be low-paid production line garment workers, who reproduce the conditions of their own exploitation by continuing to work in the industry.

Self-interest: the idea of reproducing background conditions of injustice contains the idea that the agent in question is herself bound by unjust social structures. Thus, the claim that the agent should seek to eradicate such structures can be thought of as motivated by enlightened self-interest: everyone, including the agent herself, may be better off if she does so. This is in effect an appeal to the familiar form of reasoning according to which collective action problems are viewed as iterated prisoners' dilemmas. As a criterion for determining burden sharing, self-interest makes sense: those with an interest in tackling structural injustice from a particular angle might as well be assigned the job of doing so, as they are most likely to do it effectively. As a justification for assigning responsibility in the first place, however, self-interest is problematic. For a broad class of cases in which Young wishes to claim that individuals should recognise responsibility for structural injustice, an appeal to self-interest cannot do the justificatory work. It is simply not always the case that an individual whose behaviour reproduces structural injustice will necessarily be made better off by taking positive steps to help to eliminate that injustice. Though, as just argued, consumers in the garment market do not necessarily benefit from injustice in that market, it is not the case that they are necessarily *disadvantaged*, either. In many, perhaps most cases, consumers have no self-interested motivation to disrupt the status quo. The relations between individuals reproducing structural injustice are unlike an iterated prisoners' dilemma, as cooperation may be much more costly relative to the status quo, at least for a

large class of agents.[39] Cooperation does not always pay.

Dependence: Young approvingly cites Onora O'Neill's view that one must 'accord ethical standing' in one's practical reasoning to all those 'about whom [one] make[s] implicit or explicit assumptions as a basis for [one's] own activities' (Young 2011 159, paraphrasing O'Neill 1996 Ch.4). An important formal difference between O'Neill's view and Young's social connection model is that social connection is understood as connection to *structures*, whereas O'Neill's view appeals to the idea of dependence upon on other *persons*.[40] Could dependence nevertheless provide the normative foundation for the link between social connection and responsibility? To answer this question we must look more closely at what O'Neill takes the normative role of dependence to be. 'Wherever activity is based on the assumption of others who can act and react', O'Neill writes, 'the standing of those others cannot coherently be denied' (O'Neill 103). O'Neill's aim is to establish a criterion to determine who or what has moral status, and her answer is that anyone has moral status whose agency features in our goal-directed deliberation. Thus for example, if one behaves in such a way as to conceal one's intention from others, one assumes others' agency, because one acts on the expectation others might object to one's behaviour. On Young's account of social structure, meanwhile, when one reacts to structures one does not see oneself as responding to the agency of others - rather, one encounters the effects of others' agency 'reified': one encounters them as mere background condition of one's action, rather than as social interaction. If an application of O'Neill's view to the case of structural injustice consists in the claim that it is 'incoherent' to fail to consider the agency of others

[39] As had been noted elsewhere in this thesis, Stephen Gardiner has forcefully argued that the relations between individual contributors to climate change cannot be analysed under the rubric of an iterated prisoners dilemma, as the relative benefits of climate change mitigation fall disproportionately upon people in the future rather than people in the present, and upon the economically disadvantaged rather than the economically advantaged (Gardiner 2011 25).
[40] McKeown notes this difference (McKeown 2018 492)

when acting under structural constraint, such an application cannot be apt. The reification of others' agency is supposed to be a defining feature of structural injustice.

Perhaps it will be argued that although none of these justifications can account for individual responsibility for structural injustice on their own, some combination of all of them might do the job? Perhaps, in other words, the moral justification for the responsibly of production line garment workers for structural injustice is different in kind from the justification for the responsibility of clothing company executives, and different again from that of consumers. The problem with taking this line is that the concept of "social connection" seems to drop out of the picture. If we accept a multiplicity of underlying justifications for responsibility for structural injustice, it does not appear that there is anything about social connection *per se* that makes it the case that individuals ought to be regarded as bearing responsibilities for structural injustice. What would really matter would be ability, interest, fairness, and various other considerations that happen partially to coincide with social connection. The claim that individuals who are socially connected to structural injustice bear responsibility for it would be a contingent rather than a constitutive matter, and we would need to run through every possible case of structural injustice to be sure that it was true.

In light of the fact no established form of moral justification supplies the necessarily link between social connection and responsibility, Young's argumentative strategy should be viewed, in line with Jamieson and Lichtenberg, as trying to establish a revisionary moral practice. The claim, in other words, is not that social connection generates responsibility, but that it *should.* We are faced with another instance of higher order theoretical considerations being invoked in an attempt to generate some kind of rationale for the modification of first order moral practice as it currently stands. A major strand of her argument is that the attribution of liability and the reactive attitude of blame are

inappropriate in cases of structural injustice, because these attitudes tend to isolate single individuals for criticism, implicitly absolving others. But by drawing the corollary that we should recognise forward-looking responsibility for structural injustice on the basis of social connection, she moves from considerations about the good practical effects certain practices of accountability would have, to the claim that we should adopt them. Like Sigdwick, Jamieson, and Lichtenberg, she adopts an objective stance, primarily with respect to the first personal perspective: she claims that we should each regard ourselves as accountable in previously unacknowledged respects, because to do so would have positive effects on the resolution of global structural injustice. In adopting this stance, she neglects to appreciate the internally meaningful nature of practices of accountability as they currently stand.

Is the Strawsonian Challenge Excessively Conservative or Relativist?

What we have seen, then, is it that moral revisionist responses to the problem of group-caused harms are vulnerable to a common criticism. Such arguments make the claim that certain individuals *should* be considered the objects of moral critique or approbation who are not presently considered as such. These arguments, viewed in the most general terms, rest on the premise that it would be *better* if people were judged in this new light. The problem is that this is the wrong kind of argument for this kind of conclusion. Such arguments adopt what Strawson called the objective attitude: they treat peoples' moral attitudes – their approbation of virtuous characters, their disapprobation of those who have been derelict in their duties, their attributions of accountability – as though those very judgements were themselves subject to a higher order of moral evaluation according to a criterion of goodness, and could be reinforced or thrown out accordingly. The introduction of these higher-order considerations treats first-order moral attitudes as if they were of no significance in their own right.

One response to the Strawsonian challenge might be to argue that there is nothing wrong with this application of higher-order moral principles. After all, if, say, the utilitarian principle is true, then it is quite right that our social practices should be subordinated to it, even practices so fundamental to our conception of ourselves as persons as those associated with individual responsibility. If the utilitarian principle is true, then it may be justifiable, or even required, that people be trained into better attitudes by non-rational means. Indeed, it might be argued that what we have been calling the Strawsonian view faces two important objections: that it is excessively conservative, and that it entails vicious moral relativism. These complaints arise from a misunderstanding of the Strawsonian challenge to the radical revisionist, which it is worth pausing to dispel.

Certain philosophers have argued against the validity of two related forms of methodology in ethical theory, those of moral intuitionism and reflective equilibrium.[41] Peter Singer was among the first to critique Rawls's method of reflective equilibrium, and he has continued to be vocal on that score, for at least two reasons. First, he argues the assertion that reflective equilibrium is the best methodology for moral theory implies that the truth of moral judgements can only be assessed relative to the prevailing attitudes in the societies in which they are made, making the adjudication of disagreements across societies impossible. Second, he argues the method serves to elevate mere received opinions, biases and gut reactions to the level of considered moral judgements that pretend to general validity. Instead, Singer in his more recent work defends Sigdwickian *rational* intuitionism,

[41] Moral intuitionism, in this sense, refers to the method according to which putative moral principles are tested and refined by applying them to particular cases, usually through thought experiments. It is to be distinguished from rational intuitionism, a metaethical theory which posits that there exist moral facts, which are apprehended by a perception-like faculty of rational intuition. Reflective equilibrium refers to the method, described by John Rawls, according to which our considered moral judgments are made consistent with the theoretical principles we take to govern those judgments, through a cyclical process of refining the judgments according the to the principles and refining the principles according to the judgements, until a stable state is reached.

according to which theoretical reason is used to judge that certain fundamental moral axioms are true, and the logical implications of those axioms are then upheld even when they seem to conflict with the morality of common sense.

It might be argued that Singer's concerns about reflective equilibrium also undermine the Strawsonian challenge to moral revisionism – perhaps the existing structure of our reactive attitudes is little more than a set of knee-jerk responses, reflecting engrained patterns of behaviour with no moral significance in their own right. Think, for example, of the reactive attitude of disapproval felt towards people who choose to have sex outside of marriage – in some communities, it is felt very strongly, in others, not at all. One might worry the claim that there is something wrong about bringing abstract theoretical considerations to bear on the structure of our reactive attitudes implies that we could never have grounds to criticise outdated norms, such as those pertaining to sexual purity, and that societies may be stuck with them for longer than they need be, perpetuating avoidable suffering.

This scepticism is particularly relevant in the contexts in which radical revisionism is defended. Jamieson's suggestion that 'morality has met its match in the Anthropocene' (Jamieson 2014 185) clearly echoes Singer's concern that the morality of common sense is the evolutionary and cultural artefact of an earlier age, that can no longer be expected to track moral reality in the contemporary world. A central strand of the radical revisionist position is that our intuitive moral responses developed at a time when humans lived in very small groups. Morality confined itself to the regulation of interactions between single individuals, not because those relationships were necessarily the subject matter of morality in a constitutive sense, but because the regulation of those relationships was what tended to make people's lives go better. Now the nature of our social environment has changed, meaning human actions are socially impactful in ways our ancestors would not have recognised, either because they are effective at a spatial or temporal distance, or because

they arise from the combined actions of large numbers of people. It is precisely for this reason that the radical revisionist demands novel ethical concepts.

Rawls's response to these concerns was, in *Political Liberalism*, to re-present reflective equilibrium as tool that has a place against the background of a political constructivist metaethics. On this view, reflective equilibrium is a valid way of deriving moral principles *given certain assumptions*, specifically, the fact of reasonable disagreement together with a description of what constitutes 'reasonableness'. For the constructivist, the concept of reasonableness takes over the role that the concept of moral truth plays for rational intuitionists like Sidgwick and Singer. The method produces principles that are a reasonable way of organising society given the constraints of political life. These principles may turn out to correspond to metaphysical truth, but whether they do is a question left for particular 'comprehensive moral doctrines' - Rawls's term for a particular set of beliefs about moral value that may be inconsistent with other such systems of belief (Rawls 1993 90-95).

We could describe the constructivist position in language more amenable to the rational intuitionist, to show the two approaches are not so far apart. Arguably, the fact that it is possible for reasonable people to disagree about 'comprehensive moral doctrines' is taken by Rawls to have the status of what Singer and Sidgwick would call a *moral axiom*, and the rest of Rawls's methodology can be viewed as falling out of the fundamentality of this axiom. Viewed in this light, the disagreement between Rawls and Singer is not a disagreement about whether moral principles are universal or relative, but about how fundamental we should regard certain moral axioms as being. In this way, the charge of vicious relativism can be rebutted: at the highest level of abstraction, putatively objective and universal moral principles are in indeed play, although these principles, in combination with certain contingent social facts, ground and constrain the construction of further principles which

do not pretend to absolute objectivity or universality.

The Strawsonian challenge can be seen as having force within a similar methodological context. It is not supposed to consist in, or rely upon, the assertion of moral relativism; it is not supposed to preclude the possibility a universally true moral theory might exist. Nor should it be viewed as elevating natural moral emotional responses to the level of immutable laws. Clearly, the context in which particular moral emotions are appropriate can, does and should change as we revise our theoretical beliefs about moral matters. Rather, it consists in the observation that there is something objectionable, or futile, or both, about any view that implies that theoretical reasons should be thought of as justifying the modification of our practical attitudes, even if this cannot be done by rational explanation.

To illustrate, let us grant for the sake of argument that the utilitarian principle expresses a truth, and that it follows from that truth that 'mindfulness' ought to be regarded as a virtue, as Jamieson suggests. It would not follow from these assumptions that one would be making a mistake if one failed to regard a practitioner of mindfulness as worthy of esteem. While moral theories can be a source of warrant for the appropriateness of reactive attitudes, there will always be another source of warrant, based in interpersonal relations, which firmly held belief in a moral theory does not rationally override. Even if we were so convinced of the truth of higher-order principles that we were willing to implement policies intended to modify the structure of people's reactive attitudes (something that would, it is here submitted, constitute an act of gross hubris), we should at least recognise, on pragmatic grounds, that such a policy programme would be unlikely to have a swift or lasting impact.

Moreover, Singer's two charges of relativism and conservatism dissolve when we try to give a clear statement of the critique. The claim that it should be considered a problem that a

dispute between moral cultures cannot be settled presupposes all such disputes must in principle be soluble. Take the following case by way of example. When the Shah was overthrown by forces loyal to Ayatollah Khomeini in Iran, many liberal Iranians found religious conservative values imposed upon them by force. This shift could be described as a disagreement between moral cultures (although it was of course much more than a mere disagreement). One group was acculturated to judge, for example, that it was morally wrong for a woman to fail to dress modestly, while the other was socialised to lack that judgement. What exactly is the problem here, for Singer? Perhaps it is a theoretical one: perhaps the claim is that it must be possible for one side to be determinately right and the other wrong. But the Strawsonian view does not deny this possibility. If there are moral facts about this issue, then – necessarily – of two inconsistent claims only one can be true.

Perhaps then the problem is a more practical one – it must be possible for the wronged side to attribute wrongdoing to their oppressors, it must be the case that their cause can be thought of as a righteous one, in an objective sense. But again, the Strawsonian approach does not preclude this. Victims of oppression by another culture are perfectly entitled to label their oppressors villains, whether or not their oppressors have been socialised to believe their oppressive acts righteous. If our worry is that the oppressors cannot be convinced of their own villainy, it is hard to see how it will help to explain to them that their actions are incompatible with principles that are the logical derivatives of axioms whose truth has been established by rational intuition. If the Strawsonian view implied that such disputes were less likely to be settled in favour of the right side, this would indeed be a problem. But the dispute is equally unlikely to be settled whether or not the Strawsonian approach is the right one, and whether or not moral realism is true.

The charge of conservativism, similarly, rests on the assumption that changes in our moral outlook occur because moral theorists apply the correct methodology. What exactly is the

problem with the idea that biases might be elevated to the level of considered judgments? The language of "conservatism" suggests it is specifically that we will be stuck with bad judgments for longer. Note that this is a sociological claim, it posits a correlation between the metaethical beliefs held by a given group (whether they are constructivists or rational intuitionists), and its degree of social progressiveness. It does not seem that Singer's argument is based on sociological data. In the absence of such data, we should note that the hypothesis attributes a rather self-aggrandising role to the philosopher as an engine of social progress. It seems implausible that any of the classic putative examples of moral progress, from the abolition of slavery to the normalisation of premarital sex, was precipitated by methodological changes in our metaethics.[42] If moral progress has some other driver, such as the assertion of equal status on the part of marginalised groups, whether through political campaigning or force of arms, or changes in the economic "base" giving rise to a new pattern of vested interests, then we need not fear that our metaethical views have a superficially conservative complexion.

It might be countered that it is not enough to point out that the Strawsonian view entitles one to stick to one's guns in response to moral cultures with which one radically disagrees. The implied contingency of my considered moral judgments must be significant, it might be thought. Had the Iranian liberal grown up in a different household, he might have ended up in the place of religious police officer who now berates his immodestly dressed female relative. There are at least two ways of cashing out this worry. One is a reliabilist point – if the same method (following one's considered moral judgements in reflective equilibrium)

[42] A counterexample might be the abandonment of divine command theory in favour of the theory according to which individual conscience was the source of correct ethical judgment, a change that arguably laid the ground for more liberal moral values. One response to this point might be that the shift was itself precipitated by socio-political forces, notably the Reformation, the interest Princes had in securing their power base independently of the Papacy, the rise of the merchant middle class, etc.

can lead to two inconsistent results, it cannot be a good method, or so the thought would go. Another is the idea that such radically contingent judgments would be arbitrarily ethnocentric – why is it that one should be entitled to stick to one's guns in the case of cross-cultural disagreement? What makes one's own culture superior to any another?

Both these formulations of the worry can be met with a similar answer. For a start, the fact that a method sometimes produces bad results does not necessarily make it a bad method, as long as, in those cases in which it produces good results, it does so for the right reasons. As Kelly and McGrath (2010) observe, the scientific method provides an example here: despite the fact it would produce erroneous results if the scientist were systematically exposed to unrepresentative data, it can still be considered generally conducive to valid justification. Moreover, as Amia Srinivasan (2019) has recently argued, although viewing our radically contingent beliefs as correct requires us to see ourselves as the beneficiaries of 'genealogical luck', attributing genealogical luck to ourselves is not necessarily irrational. That is to say, the Iranian liberal must regard it as a stroke of good fortune that he ended up a liberal, and therefore "right", and not a sexist religious zealot, and therefore "wrong", but nothing about this thought process need lead him to worry his beliefs are unjustified. This is because the liberal does not have to regard himself and the zealot as employing the same method. He regards his own views as arising from consideration of the salient facts, and his opponent's views as the product of some irrelevant or false considerations, or of entirely non-rational "brainwashing". Just as our everyday beliefs about physical reality are not undermined by the realisation that certain people disagree with them (for example, conspiracy theorists and practitioners of pseudoscience), neither should the justification of our moral beliefs be undermined by the existence of other moral cultures.

Quasi-Participatory Accountability as a Response to the Strawsonian Challenge

The Strawsonian view, then, should not be viewed as vulnerable to the charge that it relies

on the assertion of relativism, as it is not a claim about metaethics *per se*; rather it is a claim about how we ought to treat others in respect of their considered moral values, whatever our metaethical views happen to be. Furthermore, as we have seen, philosophical worries about relativism, and the related concern about conservatism, are in any case overblown, being based on a misunderstanding of the role of the philosopher in relation to first-order moral practice. All this can be seen to vindicate our anxiety about the moral revisionist response to the collective harm problem: we should indeed be worried that moral revisionism engages with our present moral attitudes in the wrong way, seeking to modify practices that are meaningful from an internal perspective, an exercise in hubris or futility.

Moral revisionists present us with a binary choice: either accept moral revisionism, or live with a moral framework that seems fundamentally unsuited to the kinds of harmful impacts that emerge from the uncoordinated actions of large numbers of people. If this choice really does exhaust the options, then the critique of moral revisionism just set out would be very troubling. Fortunately, there is another way forward. The system of reactive attitudes as we find it can be shown to be adequate to the task of effectively regulating the kinds of behaviours that give rise to collective harm. This can be done via the link between participatory intention and the reactive attitudes associated with the dynamic of accountability, attitudes like guilt, shame, regret, blame, indignation, pride, self-esteem, praise and approval.

As Kutz notes, the most fertile source of warrant for reactive attitudes of accountability lies in reasons of conduct (Kutz 2000 26-38). One holds an agent accountable when their conduct manifests an attitude or intention that is taken to impugn the value of the relationship that exists between one and the agent. This might be a richly textured relationship like friendship, or a minimal relationship such as the relationship that exists between members of the moral community. Which aspects of conduct might be said to

have this feature, with respect to contributions to collective harms? As intimated by Young in her objection to Kutz (discussed in the previous chapter), appeals to the idea individuals ought to be considered morally accountable on the basis of their faulty intentions may initially look unpromising in the context of collective harm and structural injustice. Contributing to collective harm cannot realistically be regarded as part of the content of someone's intention when they engage in carbon-intensive behaviours, nor can reproducing structural injustice be considered an intentional result of purchasing clothing.

Despite these obstacles, one's actual intentions and attitudes, viewed in the correct light, will be seen to be appropriate objects of moral opprobrium in just the kinds of cases that revisionists claim call for the application of novel values. The attitudes in question are attitudes according to which we regard ourselves as participants in certain kinds of groups. As intimated in the previous chapter, Christopher Kutz has offered a preliminary description of the sorts of attitudes we are looking for, but his account lacks cogency in certain respects and is in need of further elucidation. This is the task to which we now turn. Kutz writes:

> *Consider tailpipe emissions, especially in the United States, from large and inefficient automobiles. The taste for such automobiles was a product of many factors, including low fuel prices; but also of socially reinforced trends of admiring (rather than disdaining) large SUVs; collective action effects of individuals feeling that they were endangered by others' large cars unless they too bought large cars; and the general disinhibiting effect of seeing others in such cars as well. ... No one individual's choice of what car to drive probably made a difference to anyone else's behaviour, taken on its own – but overall a network of collective choices was built up out of these individual interactions. And so, in such a case, we can say that the global increase of CO_2 emissions is attributable to the collective, and not merely*

parallel, acts of US consumers. (Kutz 2015 359-360)

The implication drawn is that individuals share in collective guilt for the harms perpetrated by the group of SUV drivers in the US and are on the hook to share in reparative duties arising from such harm as a result. Guilt is here regarded as a reactive attitude rather than just an abstract ascription of fault: it is a warranted emotional response. Kutz is clear that participatory intention is a necessary condition of collective action; therefore, it is puzzling that this description does not obviously mention participatory intention. Plausibly, the thought is that the attitudes described serve as a kind of proto-participatory intention or implied participatory intention. Through the attitudes mentioned, individuals 'jointly constitute a normative system that guides individual choice' (Ibid.).

The account may appear similar to Young's social connection model of responsibility, which, as we saw earlier, is best understood as linking responsibility for structural injustice to acts that reproduce the choice conditions that both constitute and give rise to structural injustice. There is, however, an important distinction between the two. Young assigns responsibility to individuals who reproduce constraints that are experienced *reified*, as impassive background conditions. Kutz, for his part, points to the creation of a *normative* system. He thereby highlights the participatory nature of the sorts of roles in question – when one purchases an SUV, one intends to affirm their value *to and with* other agents. There is a prescriptive element to such a choice: it re-establishes and sustains the general norm according to which the object of choice is choiceworthy. In this sense, such agents share an intention to do something together, namely to constitute this very normative system by which they each then are bound.

As alluded to in the previous chapter, a potential problem with this approach is that it seems to rely on a supposed link between guilt and affirmation - psychological affects which,

it might be thought, do not sit naturally together. This is effectively the problem Kutz himself notes when he observes people may respond to the attribution of collective fault by deciding that they should band together more closely, 'to protect [their] shared way of life' (Kutz 2000 187). Any social practice one affirms, especially in the normative-prescriptive manner just described, could not apparently be a practice that one reproached oneself for joining. This objection is can be countered, however: the relationship between affirmation and guilt is not static, but dynamic. First, one participates with others in jointly constituting the normative system that leads to the popularity of SUVs, then, recognising the demonstrable harm and suffering that arises from the effects of that system taken as a whole, one regrets one's participation. Moreover, one is then motivated to help to repair the damage the group has caused, precisely in order to express one's disavowal of one's former association with it.

Note also that one's motivation here is not dependent on one's having foreseen that the system in which one was participating would have harmful effects. One should feel what Bernard Williams called 'agent-regret' (Williams 1981). Even though, with the information one had at the time, one knew no better than to participate, and so perhaps could not be expected to have acted differently, one should still reproach oneself for having been involved at all. As Williams notes, 'we feel some doubt' about an agent who too lightly exonerates herself for the bad results of her intentional actions, even when those results are accidental. Such a person fails to take the strains of her own agency seriously. Although Williams himself examines cases of agent-regret where there is a direct causal link between the individual agent and the regretful consequence, there is no reason why agent-regret should not be felt for the consequences of collective acts just as for individual ones.

Another problem with quasi-participatory intention as a source of accountability for individual contribution to collective harms is that there remain many instances of

contribution in which individuals cannot obviously be described as having participatory intentions of any kind, whether intentions to constitute a normative system or otherwise. A mother who drives her child to school every day may do so solely because she wants her child to get a good education, there are no schools in cycling distance, and she cannot afford to move into the neighbourhood of a suitable school. She does not apparently intend that any norms be reproduced; indeed, her situation seems much better theorised in Young's terms, as bounded by reified objective constraints, rather than an intersubjective normative system.

A quick response is available, though some might feel it concedes it too much ground to the moral revisionist. Perhaps individuals like the mother are just not appropriate subjects for guilt, agent-regret or reparative duties. Perhaps it is simply true that no one is responsible for the harm caused by the aggregate emissions of many such individuals, or perhaps the best we can do is assign responsibility for their emissions at the level of the nation state. Some may worry that the moral revisionist could use this response to trap us between the horns of their dilemma: it simply cannot be the case, the revisionist may argue, that such a significant quantity of human-caused suffering is morally unaccounted for, suggesting once more that our moral concepts are indeed unfit for purpose and in need of reformation. That said, others might find the response satisfactory; it may be the best we can do.

Another possible response has its own difficulties. The mother, we might observe, clearly does embody some participatory norms in her choices – she affirms and reproduces the norm whereby parents and guardians value ensuring their children receive a decent education. Here again we are troubled by the incongruity between affirmation and guilt – how could the mother affirm her participation in the reproduction of the norms of good parenting, and yet regard that very behaviour as engendering guilt? Surely, to feel guilt

would be to disavow the rightness of her actions, something she would doubtless be practically unable to do. But perhaps there are separable elements in this story. Of course, it is not her child's wellbeing that she disavows, but perhaps her own individualism, her failure to consider that what was best for her family might have broader social ramifications, of which she had failed to take stock. One appropriate response to this self-reproach might be to make amends for the bad consequences of her participation in individualistic norms in other areas of her life.

The Risks of Revisionism

Several aspects of humanity's relationship with the highly interconnected, global societies which it now inhabits suggest, to some authors at least, that we need totally new moral concepts, our old ones beings unfit to deal with phenomena like New Harms, structural injustice and environmental virtue in a 'dark time'. What we have seen is that this call to revisionism comes with its own serious drawback: it is at odds with a very appealing way of conceiving of the functioning of moral phenomena like accountability, desert, condemnation and justice, an approach that we have called the Strawsonian view. On this view, the appropriateness of certain moral emotions like esteem or indignation is determined by the interaction between our natural dispositions and the network of social relationships in which we find ourselves. Radical moral revisionism, it has been argued, essentially implies the idea that this system of emotional responses is in some sense wrong and should be modified.

Unfortunately, because the system of reactive attitudes operates independently from our theoretical beliefs about morality, it does not seem such a modification could be effected except by extremely distasteful, impractical and uncertain means, bypassing our reason with programmes of propaganda, conditioning, or other such schemes. Fortunately, we have been able to show that our existing moral concepts are not so poorly fitted to the modern

world as the revisionists claim, making their flawed project unnecessary. Individuals can be helped to recognise their own accountability for their participation in collective harms, by drawing their attention to their intention to participate in certain loosely collective endeavours with harmful consequences, engendering complicitous accountability. An example of such an endeavour was reproducing the norms that make SUV driving attractive and popular, although other such collective activities could be cited. The reactive attitude of collective guilt that goes along with the recognition of complicitous accountability should cause us to recognise a duty to take on burdens associated with the repair of global collective harms such as climate change.

The next chapter will address a final potential source of reasons for individuals to refrain from contribution to collective harms: considerations of hypocrisy. Such arguments are significant in activist discourse and in the media at large, but have received relatively little attention in the climate ethics literature. The claim considered is that norms related to the avoidance of hypocrisy provide people committed to combatting climate change with special reasons to reduce their individual emissions. It is both intuitively plausible and borne out empirically that the perception climate advocates are failing to reduce their own emissions makes their advocacy less effective (Attari et al. 2019). The proposal is therefore that it is a worthwhile endevour to consider whether hypocrisy avoidance can be said to rise to the level of an individual moral duty. While such duties would not provide a case for individual outcome responsibility for climate change as such, they would go some way towards solving the collective harm problem viewed as a paradox of practical reasoning: they would help to justify the sense of a duty to reduce emissions that many individuals already acknowledge. As we shall see, however, the range of cases in which hypocrisy avoidance should be of serious moral concern is very small.

6. The Morality of Hypocrisy in Climate Action

It is now a widespread view, both in the academic community and in the wider public sphere of campaigners, authors and journalists, that the narrative which illustrates our impacts and responsibilities with respect to climate change in terms of an individual "carbon footprint" is seriously misguided, perhaps systematically so. Journalist David Wallace-Wells writes, '[a]lmost as a prophylactic against climate guilt, as the news from science has grown bleaker, western liberals have comforted themselves by contorting their own consumption patterns into performances of moral or environmental purity – less beef, more Teslas, fewer transatlantic flights. But the climate calculus is such that individual lifestyle choices do not add up to much, unless they are scaled by politics' (Wallace-Wells 2019). Individual mitigation efforts, the thought goes, are a kind of essentially selfish displacement activity, through which we seek, ineffectually, to wash our hands of involvement in the crisis. Environmental advocate Mary Annaise Heglar, of the Natural Resources Defence Council, agrees on the symptom, but not the diagnosis: '[t]he dominant narrative around climate change tells us that it's our fault. We left the lights on too long, didn't close the refrigerator door, and didn't recycle our paper. I'm here to tell you that is bullshit...The Oil and Gas Industry is gaslighting you' (Heglar 2018).

Where Wallace-Wells attributes our concern for personal emissions reductions to the psychological incapacity of each of us to face up to the enormity of the problem ("There must be *something* I can do"), Heglar detects more sinister forces at work behind public discourse: the tendency to focus on demand-side changes exists *because* it serves the interest of those on the supply side. If Heglar is right, then one of the key weapons of psychological warfare, of this 'gaslighting' of people with environmental concerns, must surely be the charge of hypocrisy. Examples of this rhetorical move are just as easy to find.

Here is Julie Kelly, contributor at *The Hill*: 'being a climate change believer means never having to say you're sorry, or at least never making any major sacrifice to your lifestyle that would mitigate the pending doom you are so preoccupied with (but, sea ice!). You can go along with climate change dogma and do virtually nothing about it except recycle your newspapers while self-righteously calling the other side names' (Kelly 2016). Kelly demonstrates that, as a mode of criticism, the charge of hypocrisy is as convenient as it is powerful: it permits its wielder to shame their opponent for failing to do enough for their cause, without committing to a view on the validity of the cause itself. It turns an opponent's own weapons upon her, all the while hiding in a place of safety. The climate change believer, it is suggested, criticises others while doing nothing herself. The climate change denier meanwhile, who presumably thinks none of us have any reason to change our behaviour in any case, nevertheless traps the believer between the charge of failing to live up to her own standards and the imputation that she herself does not truly believe.

Clearly, then, the charge of hypocrisy presents a strategic threat to the objectives of climate change advocacy, and those engaged in this field must develop effective responses in the practical pursuit of the goal of tackling global environmental degradation. But as we prepare our lines of defence, redirecting the conversation back onto the right track of promoting coordinated political action, we may perhaps find ourselves wondering whether we deserve to feel the sting of the anti-hypocrite's barbs. Should we be worried about our hypocrisy, not only at a tactical level, but at a moral level as well? It has been suggested that committed environmentalists have a particular obligation to reduce their personal carbon footprint, grounded in 'an obligation to avoid hypocrisy' (Hourdequin 2010 448). But what exactly is wrong with hypocrisy, in the environmental context? Should we regard it as a form of wrongdoing at all?

Systematic Hypocrisy

Concern for hypocrisy has historically been particularly salient in a religious setting. The Evangelist Matthew depicts Jesus admonishing the hypocrite first to cast out the beam from his own eye before he attends to the mote in his brother's. In the same sermon, Jesus warns the assembled congregation to beware false prophets, as they may be wolves in sheep's clothing. Should the latter directive inform our interpretation of the former? In other words, are we to avoid hypocrisy in our dealings with our brothers, in case they take us for wolves in sheep's clothing, and the impact of our moral message is diminished? On this reading, it would appear that even the biblical Jesus viewed hypocrisy more as a strategic concern than a sin or fault in its own right. Judith Shklar (1979) depicts hypocrisy and puritanism locked in a kind of a vicious dialectic: the religious requirement of faith encourages an exaggerated pretence of religiosity, which gives rise to stricter demands for sincerity, followed by even more ostentatious displays of faith. That the language of 'dogma' and 'purity' lingers on in the contemporary journalistic sources cited above may be seen to confirm the theory of an essential connection between anti-hypocrisy and the unmasking and ridicule of puritanical attitudes. These observations lay out the terrain, but do not yet explain, vindicate or indeed debunk our concern for the avoidance of hypocrisy. As Shklar notes, 'to fail in one's own aspirations is not hypocrisy' (Ibid. 5). Puritans were often engaged in a private internal struggle to repress in themselves the same supposedly sinful urges they were wont to condemn in others, often at the expense of their own psychological wellbeing. If such people are futilely consigning themselves to a cycle of self-loathing, then they are to be pitied, not reproached.

It seems therefore that the traditional religious critique of hypocrisy simply targets the lack of sincerity, which is to be feared for either of two reasons. For one, it provides a cover for the secretly faithless, who may present a danger for the simple reason that they are judged

to be prone to other forms of harmful social behaviour. For the other, the puritans' inability to abide by their own constraints reveals those constraints to be unwarranted, or at best supererogatory; this species of hypocrite pretends to 'a piety greater than God requires...a covert form of pride', as the Pauline dictum would have it (Ibid. 4). If this is right, we have to conclude that hypocrisy that neither masks immoral designs nor aims at unjustified self-promotion is not to be especially resented or condemned. And indeed Shklar concurs: she praises Charles Dickens for 'never forgetting the difference between wickedness and mere pretention', even as he excoriates the 'humbug that sugarcoats meanness' (Ibid. 7). While we are revolted by Uriah Heep's affected humility, Dickens's target here is rather a social critique of the system of class hierarchy and patronage wherein Heep has correctly identified fawning humbleness as a means of gaining advancement.

Why, then does hypocrisy seem to have more moral significance than mere insincerity, and why do accusations of hypocrisy carry so much more rhetorical force than anything so easily brushed off? Shklar's thought is that a change in our moral circumstances allowed hypocrisy and anti-hypocrisy to become a 'discrete system' within moral discourse. This explains the peculiar power anti-hypocrisy gives political opponents to 'wound without altering one another' (Ibid. 11), just as Kelly is able to land an attack on environmentalists without having to defend her own view. The change in question is the collapse of our shared belief that morality is exogenously given – by God, typically – meaning this public moral code is no longer common ground between disputants. Without a shared stock of moral knowledge, hypocrisy discourse become a battle for the legitimacy of sources of moral authority: individual conscience on the one hand, and social convention on the other. By attacking the strength of their opponent's professed convictions, each side can shake the other's psychological dependence on their favoured source of moral authority, and thereby cause 'psychic annihilation' (Ibid. 12). This is illustrated in the Victorian conflict between

traditional family values and sexual libertinage: the monogamist is genuinely disturbed by the libertine's accusation that he is denying himself the chance to experience true love – something he himself claims to prize more than mere hedonism. The libertine is similarly shaken by the suggestion that he does not really derive satisfaction from fleeting trysts and is denying himself the higher pleasure of monogamy.

Anti-hypocrisy is thus a particularly effective weapon in ideologically contested territory. When the right-wing press attacks a Labour politician for sending her children to private school, the imputation is that socialist opposition to the institution of private education is therefore unfounded. More precisely, it counts as evidence that the politician herself does not trust her own convictions. The problematic cases are those in which, because they take place in a domain of discourse pervaded by 'essentially contested concepts' (Gallie 1955), disputants are making their cases from very insecure positions, upon which sincerity is one of very few means of anchoring oneself, and the charge of hypocrisy one of few means of launching an attack. A British Conservative MP recently lambasted the left for lining up to condemn US President Donald Trump's state visit to the UK, while a few years previously, during the State visit of Chinese President Xi, they had supposedly remained silent about racist human rights abuses committed by his regime. This attack had some force, until it was pointed out that as the MP was delivering it to camera for broadcast on television news, the flag of Apartheid South Africa could be seen proudly displayed on the mantelpiece of his parliamentary office in shot behind him – as Shklar observed, through this form of systematic, ideological hypocrisy, 'politics becomes a treadmill of dissimulation and unmasking' (Ibid. 13), a cycle whereby disputants take it in turns to disarm their opponents without ever engaging in substantive argument. Systematic hypocrisy should not be confused with 'naïve hypocrisy', or representing oneself as especially virtuous to mask indisputable wrongdoing, which is rightly regarded as a symptom of tyranny. In such cases,

hypocrisy is not our central moral concern, rather it is the initial wrongdoing itself. Confronted with the decree that some animals are more equal than others, we resent first the injustice of preferential treatment for the chosen few, and the false mask of egalitarianism only in a derivative sense. Deception merely compounds the wrong by making it less likely to be righted.

As the old show-business joke has it: "sincerity is everything – learn to fake that and you've got it made". At this point, we may be inclined to agree: according to the analysis thus far, drawn from Shklar, the only concern the environmentalist should have is to avoid the *appearance* of hypocrisy. If a certain degree of dissimulation is necessary to achieve the aim of avoiding climate catastrophe, then the fault is not to be laid at the door of activists, but upon a political reality in which opponents are forever eager to cry conspiracy, calling environmentalists' motives into doubt simply because it is politically expedient for them to do so. Insofar as the charge of hypocrisy successfully wounds, it is symptomatic of the fact certain questions thrown up by climate politics are very difficult, and it is hard to be assured in our response. But for the most part, the attacks of climate change sceptics are attempted in such general terms that they are unlikely significantly to succeed in destabilising environmentalist convictions. Indeed, that the libertarian right are so very fond of characterising environmentalists as hypocrites, while environmentalists remain for the most part unembarrassed by the charge, can arguably be explained by the difference in the degree to which each side regards the discourse as fundamentally ideological. While admittedly no evidence is here offered for this assertion, it is anecdotally plausible that most environmentalists regard their beliefs about the need for swift political action to combat climate change to be founded on very uncontroversial moral principles about the importance of avoiding catastrophic harm, combined with propositions they regard as scientifically settled and not the proper subject of political debate. Climate change sceptics,

meanwhile, regard those propositions as a power-grab that science is attempting to make in the domain of politics, and therefore go after them with the political weapon of anti-hypocrisy, often to their opponents' bemusement.

One of the greatest insights that can be drawn from Shklar's classic essay is that the most important task is not to define the wrong of hypocrisy, but to identify whether any of those phenomena that go by the name of hypocrisy can properly be considered morally objectionable. In the spirit of this project, it is arguable that there remains a conception of hypocrisy that lies somewhere between Shklar's distinction between naïve hypocrisy and systematic hypocrisy, whose moral significance has been neglected. It is already suggested, perhaps, in the link drawn between hypocrisy and tyranny. There is something contingently tyrannical about the tyrant's tendency to conceal evil deeds beneath a mask of virtue. This is clearly a useful tactic that many tyrants must surely fall into adopting. But there is something *constitutively* tyrannical about making prescriptions concerning the behaviour of others, whether moral or legal in character, without applying those prescriptions to oneself or one's favoured few. Compare the hypocrisy of the Labour politician discussed above, with another politician, who bans private education and extols the virtues of the state education system, over which she has full responsibility, while secretly sending her own children to private schools abroad. This is not naïve hypocrisy: it is not the case that the political programme being pursued by this politician is inherently evil or wrong.[43] The education system, we can stipulate, is not so bad that it can be regarded as a dereliction of the state's duty to provide adequate education, if the state has such a duty. Nevertheless, it

[43] Perhaps some readers would assert the banning of private education is indeed an evil. It is hoped such readers will agree that it should in any case be possible to construct a parallel example they find less controversial.

is more than systematic hypocrisy, insofar as our reasons to condemn the individual hypocrite seem to run deeper.

This form of hypocrisy is essentially political: its concern is hypocrisy in the exercise of power. It is a political cliché often repeated that laws are like sausages, in that they cease to inspire respect in proportion as we know how they are made. Indeed, it is sometimes suggested that the practice of politics is inseparable from hypocrisy, and that politics cannot effectively be sustained without it. It is good that we have higher standards in the public sphere than in the private sphere, as in this way we reproduce the values and aspirations that constitute our political culture. It is better that politicians maintain the dignity of their office by behaving as though they are worthy of it, than that they openly flaunt their moral inadequacies in a way that would delegitimise political institutions. As Shklar noted, many expressions of racism have now been banished from public life, although they undoubtedly persist in private. Rather than condemning the insincerity of this discrepancy between our political and private faces, however, few egalitarians would deny that we should be gratified by this modicum of progress, so long as we do not rest on our laurels.

'Second-order' Hypocrisy

David Runciman (2009) agrees that while hypocrisy is endemic in politics, it is largely harmless, and even helpful, with the exception of what he calls 'second-order hypocrisy'. This, it would seem, is to say a kind of puritanical avowal of ideological purity, designed to set oneself above one's equally two-faced and corruptible political contemporaries, winning public trust by conducting witch-hunts against double-dealing, defenestrating those who, while perhaps not quite innocent, are at least no worse than anyone else, and distracting public attention from the much greater political faults of mere incompetence and misguided policymaking. Second-order hypocrisy is hypocrisy in one's professed assessment of the political system in which one finds oneself: if democracy is indeed reliant

on hypocrisy in order to sustain itself, then the second-order hypocrite is one who claims to be able to cast out hypocrisy from public life, wilfully ignoring the dangers of doing so. The tyrant is perhaps an example of such a second-order hypocrite: Robespierre and Stalin attempted to sustain revolutionary fervour by rooting out enemies of the people, fifth columns, the rear guard of counter-revolutionary class consciousness, degenerates incapable of re-education. This second-order hypocrisy was particularly dangerous because its pretence the project could one day be completed rendered it inherently unstable, relying as it did on the continual devouring of its own children.

Runciman's second-order hypocrisy is somewhat under-described. Key to its distinctiveness seems to be the observation that hypocrisy itself is socially useful. If this is correct, then the second-order hypocrite would be at fault simply because he destroys something of value. Whether he himself believes his anti-hypocritical invective would seem to be irrelevant with respect to the degree to which we resent him (especially considering that anyone who resented him would already have to have been convinced of hypocrisy's usefulness). One of Runciman's examples may help to elucidate matters, and it also brings us closer to the specific domain of political discourse in which we are particularly interested. First-order hypocrisy, Runciman suggests, was exemplified by Al Gore, for many years the figurehead for climate change advocacy in US public life, who, it emerged, had a particularly large carbon footprint, with his home's energy consumption being much higher than those of his neighbours. Second-order hypocrisy, meanwhile, was exemplified by then-Tory leader and future UK Prime Minister David Cameron, who would allow himself to be photographed cycling to work in an obviously calculated performance of his green credentials, while instructing his chauffeur, carrying his shoes and briefcase, to follow at a discreet distance. Cameron's behaviour was the more 'corrosive', Runciman argues, because it 'makes a mockery of the whole business of public enactment' (Ibid. 224).

While it is tempting to agree with Runciman that Cameron strikes one as the more contemptable of the two, it is not immediately easy to see why Cameron's case should describe a form of hypocrisy of special salience. The key difference, it would appear, is that - even if we assume for the sake of argument that Gore is indeed doing the very thing he cautioned others against doing - his behaviour might at least be put down to akrasia or moral weakness.[44] Cameron's, meanwhile, is shallowly performative. He manipulates others by claiming commitment to a particular popular cause, while the circumstances reveal that commitment to be entirely absent. Cameron's, then, is indeed a more clear-cut case of social threat: we allow no possibility that he is a pitiable self-hating puritan, and every possibility that he is a wolf in sheep's clothing. The situation is analogous to the threat that the early Christians saw in the hypocritical believer: that they would use the pretence of faith to gain trust, giving them the chance to abuse it. In the modern context, it seems if anything too generous to call this hypocrisy – it is closer to fraud. Cameron wanted to give the electorate reason to believe he could be trusted to enact green policies, when in fact he could not - or at least, the evidence he presented to the electorate as proof of his good intentions was deliberately falsified. Similarly, Runciman argues that Hobbes in *Behemoth* took a particular crime of the parliamentarians to have been concealing their intention to 'challenge the sovereignty' (Hobbes 1839b 197) until after King Charles had been executed, and thereby seizing power under false pretences.[45] If this interpretation is correct,

[44] The assumption that Gore is an unmitigated hypocrite is charitable to his detractors. Gore has other defences available to him – that he never specifically demanded individuals lower their carbon footprints unilaterally, that his advocacy work has (perhaps) made a great deal of positive impact, and his personal contribution very little negative impact. He may also have purchased carbon offset. But for present purposes, let us try to see him through the eyes of the anti-hypocrite.

[45] Runciman's characterisation of the Hobbesian complaint is more complex than this. The complaint seems to be that Parliamentarians used the demand for religious liberty as a stalking horse, concealing their true intention: the seizure of power. The hypocrisy consists in the fact that they misrepresented the nature of sovereignty, on Hobbes's account: the King could not grant freedom of conscience, because no sovereign entity could grant freedom of conscience without dividing its power. This was equally true of Parliament and the Lord Protector, as was manifest in the persecution of Catholics during the interregnum. The hypocrisy of Parliament was therefore to

then the distinction between second-order hypocrisy and everyday political hypocrisy is that many politicians really can be expected to pursue their public political agenda, even though the pursuit of that agenda will sometimes lead them to adopt positions which appear to be inconsistent with positions they have adopted in the past, or with elements of their private lives. The Labour former Deputy Prime Minister John Prescott was often mocked in the press for owning two luxury Jaguar cars; later, Labour Leader Ed Miliband received the same treatment, being pilloried for having two kitchens in his house. But while these jibes might be briefly embarrassing, there is nothing that actually precludes a relatively rich politician from being on the side of the poor, just as there is nothing that stops one with baroque sexual tastes from being socially conservative, or nothing that prevents a draft-dodger from being responsive to the wishes of the military establishment.

Runciman, then, can be read as saying that while managing one's self-presentation for political purposes is unobjectionable, conning one's way into power is not. Wearing a particular mask, suited to one's legitimate political role, is necessary – as Shklar recognised through the example of Benjamin Franklin – swapping masks arbitrarily for narrow tactical reasons rightly elicits condemnation. The Great Terrors of France and Russia can be seen to fit this pattern as they are characterised by the wrong of maintaining power through a pretence of political purity (compounded of course, by the much greater wrong of extensive unjustified killing and repression). Where in a democracy, the mask of political purity is

pretend a concession no sovereign could grant was a condition of the King's legitimacy, potentially undermining the legitimacy of the very institution of sovereign power. But given that Runciman implies that the Parliamentarians' hypocrisy is of a common kind with Cameron's, it is nevertheless most cogent to characterise Runciman's second-order hypocrisy as a species of fraud, specifically fraud regarding the limitations of political reality. Runciman's project is not to identify the *wrong* of hypocrisy, but to identify a form of hypocrisy that is a genuine *political* vice, something that ought to be banished from public life. Thus although the form of hypocrisy he identifies is arguably a form of fraud, that is not to attribute to Runciman the view that it is wrong because deception is wrong. Rather, the claim is that it is politically dangerous because it is a threat to norms and values that ground political stability.

used to seduce the voter, under tyranny it is used to manufacture a more tenuous form of consent: fearful submission.

Runciman's second-order hypocrisy is perhaps a sub-set of the constitutively tyrannical political hypocrisy whose description we are beginning to approach. The second-order hypocrite makes an exception of herself because, if her project of political purification (draining the swamp?) were ever completed, she too would have to face the firing squad. But the similarity between the two cannot be drawn too closely. There is something distinctly contingent about the wrongness of second-order hypocrisy: it lies not so much in the relationship between the hypocrite and the one who resents her, but in the hypocrite's tendency to destabilise the social order. The wrongfully hypocritical sovereign misrepresents constitutional reality. If, for example, he claims legitimate authority to flow not from himself, but from God, in order to wreathe himself in piety, he divides Sovereignty by ceding power to the Church, potentially undermining the core Hobbesian political aim of security (see Hobbes 1839a 693-697, cf. Runciman 2009 37).[46] Cameron similarly subverts the norms of the public sphere by playing a role that is not his, undermining our expectation that his public persona will at least remain consistent, even if it does not reflect his "inner self". This makes it harder to have faith in democratic politics. Had Runciman been writing ten years later, it seems certain that Donald Trump would have been a supreme example of the second-order hypocrite, a politician who seems hardly to care what he says, let alone whether it is true, so long as it is something he imagines "his people" want to hear (hence his close association with the coinage "post-truth politics"). In sum, the

[46] Hobbes makes this point indirectly, by praising the wisdom of the ancients who organised religion such that obedience to the civil laws was pleasing to the gods, rather than making the gods the source of civil authority.

concept of second-order hypocrisy essentially brings problematic hypocrisy under the more general categories of political chicanery and "dirty tricks".

In the politics of climate change, there exists a peculiar example of a related kind of anti-hypocritical discourse identified in (Gunster et al. 2018), which the authors refer to as an 'institutional cynicism' discourse. The term describes a discursive approach which accuses politicians of hypocrisy in failing to live up to their international commitments, but (perhaps counter-intuitively) this accusation is then mobilised as part of a conservative, climate-sceptical narrative. The strategy is to suggest that politicians' failure to live up to their commitments demonstrates that these politicians are secretly opposed to climate action, but have been forced to pay lip-service to the project of emissions reduction by a dominant though misguided international elite. Strangely, climate change activists and sceptics may indeed come together in their assessment of politicians' motives, but disagree on the implied conclusion – that hypocrisy is an inevitable, even smart response to a political reality that forces politicians to accept impossible targets. The cognoscenti, the underground community of climate sceptics, are then invited to continue to support conservative politicians even if they publicly feign concern about climate change, and the public are dissuaded from drawing any inference from the number of politicians discussing climate change to the reality and seriousness of the problem.

Here, it is not so much the hypocrisy, as the vindicatory discourse surrounding it which is 'corrosive', in Runciman's terms. If conservative pundits are correct that such politicians do not intend, by striking a conciliatory pose with climate activists, to deceive voters, but rather - with a nod and a wink - to bring voters along with them, then they are not second-order hypocrites, but mere first-order hypocrites, in Runciman's sense (and if the failure to tackle climate change is an evil, then they are naïve hypocrites, in Shklar's sense). It would be an odd kind of double-counting of wrongs to say that such politicians do wrong by their

hypocrisy, insofar as they provide other commentators with an opportunity to interpret their actions in a way that corrodes the public understanding of political reality. And while these political commentators are guilty of a similar kind of recklessness with the truth as Runciman's second-order hypocrite, this recklessness is not itself hypocrisy. Thus the discourse of institutional cynicism, while interesting, does not contain a variety of morally problematic hypocrisy distinct from the ones already described.

Hypocritical Moral Criticism: A Wrongful Form of Hypocrisy?

The thesis we are pursuing is that there is a political form of hypocrisy which can be considered an inherent wrong, independently of whether it has destabilising effects on the social order. Though Runciman provides us with an account of an inherently political form of hypocrisy, he does not provide us with a model for a form of hypocrisy that constitutes a moral fault in of itself. Hobbes's hypocritical sovereign supposedly commits injustice, in that he fails in his primary political duty by dividing his powers, but this claim depends on Hobbes' idiosyncratic conception of justice, which we need not accept. It seems odd, on Runciman's account, to say that David Cameron commits injustice, or wrongs anyone.

As R. Jay Wallace writes, 'moral values typically have an interpersonal dimension, a connection with objections or complaints that could be brought by, or on behalf of others' (Wallace 2010 313); the concern for second-order hypocrisy is in that sense rather a principle of political prudence than a moral principle in its own right. Wallace's own theory of morally wrongful hypocrisy brings us somewhat closer to the thesis we are pursuing, locating it as it does in hypocritical *moral criticism*, as when one expresses resentment towards another for deceiving her, when one has oneself frequently deceived the other in the past. To blame someone, or to resent someone, is to be subject to a particular reactive attitude – an emotional response. One resents someone when one takes her to have undermined something one values. Moral indignation is a reaction against another for

having flouted the values that morality embodies, values that are important to one – such as relationships of respect. Blaming, therefore, on this view implies a certain kind of commitment: it is a manifestation of commitment to the value of treating others as morality demands. When one blames another, say for lying to one, and the other points out that one has frequently lied to her in the past, this seems to generate a strong pressure to apologise. Why? On Wallace's account, it is not because of some rational pressure of consistency – why should inconsistency be a *moral* failing? Rather, it is because, if one continues to reproach the other without acknowledging fault, one is granting a higher moral status to oneself than to the other. It is as if one is saying, "Certainly I lied to *you*, but that is quite different from *your* lying to *me*".

More needs to be said – why exactly does a failure to acknowledge the wrong of lying in oneself indicate that by reproaching those who lie to one, one fails in one's commitment to morality? The thought is that a commitment to morality contains a principle of equal moral consideration of persons. The practice of blaming can be viewed as a practice through which the burden of opprobrium is distributed. Wallace's thought seems to be that an equal distribution of opprobrium for the same offence 'operationalize[s] an attachment of equal significance to [people's] basic interests' (Ibid. 333). Blaming another while failing to critically scrutinise the same behaviour in oneself is supposed to indicate that one attaches greater significance to one's own interests than those of the other, because we all have an interest in protection from opprobrium. Thus he explains the force of the biblical metaphor of the mote and the beam: morally criticising others carries with it a commitment to scrutinise oneself. 'There is something unseemly', Wallace writes, 'about resentment and indignation of others if they do not go along with a willingness to acknowledge publicly your own moral shortcomings' (Ibid. 337). Wallace, therefore, echoes St Paul, for whom hypocrisy is 'a covert form of pride', if by 'pride' we mean

systematically evaluating oneself more highly than others. What remains somewhat unclear is why hypocritical moral address should necessarily or constitutively involve such an inegalitarian evaluation, as Wallace would have it. Why should one be making any evaluation at all? Perhaps one really is just being inconsistent, and there is nothing more to say. The point is especially clear when one's hypocrisy does not involve making an exception of one's own misdeeds, but those of a third party. If A criticises B for lying to C, but fails to criticise D for lying to E, B would seem on Wallace's account to have a complaint against A, insofar as she evaluates B's interests less highly than D's. But this looks like a classic case of "two wrongs don't make a right" – it is the indignation of the pupil who receives the brunt of the teacher's reproval because she happened to be the one who was still talking when the teacher turned around. It is not obvious that A lacks the standing to criticise B until she has apportioned equal blame for all comparable cases.

Wallace can defend himself in the following way: suppose that A and D are both men, whereas B is a woman. Now it looks like we have a much stronger case for the wrongness of A's hypocrisy, because it is a symptom of systematic prejudice. Wallace is comfortable with the idea that this seems to blur the wrongs of hypocrisy and prejudice into one: he takes it to be evidence for his conception of hypocrisy that it elucidates a less visible connection with another form of wrongdoing. It is a feature of Wallace's account, though, that hypocritical criticism can only be considered wrong if the hypocrite is given a chance to reflect – it imagines a dialogic interaction between addresser and addressee. It also depends on a sharp distinction between moral criticism and moral advice. It is not obvious that this distinction holds up. The case of the teacher is telling here: very often, people have standing to issue moral criticism in virtue of a particular role they hold. To appeal to another schoolyard case, pupils feel the sting of hypocrisy when a teacher reprimands them for smoking, knowing very well she is off for a cigarette herself. But this hypocrisy is

certainly not morally problematic: the teacher has standing, and even a duty, to issue a reprimand, because it is her role to uphold a set of rules. Politics and the judicial system furnish similar examples: a Conservative party whip recently (at time of writing) ordered his colleagues to vote with the government, and then failed to vote with the government himself. Whatever the truth about whether he really intended to indicate his lack of assent with the motion, we would not say that he wronged any of his colleagues by criticising them for contemplating disloyalty, as his role gave him standing to do so. Similarly, while it is of course less than ideal for a judge to have a criminal background, she does not wrong criminals by reproving them for offenses she herself secretly commits. Indeed, we might say, it is better that at least one of them faces justice, as long as the judge passes sentence in accordance with established norms, and does not issue excessively harsh sentences in order to hide her own crimes. If she were excessively harsh, to be clear, she would be blameworthy for passing an unjust sentence, and for her crimes, but there is no reason to think her hypocrisy itself would be wrong.

The contention is that these cases are not exceptional. The standing to uphold morality, it is reasonable to suppose, is not limited to particular offices or roles; these cases simply help to elucidate the moral relations in play. Wallace's concern that hypocritical criticism manifests a failure to view people as equals is real, but only if the hypocrite is recalcitrant in the face of reflection. If one is unaware of the fact that one is making an exception of oneself or one's friends, hypocritical moral criticism does not necessarily involve anti-egalitarian attitudes – it could be mere inconsistency. Even if one is aware, criticism may still not involve anti-egalitarian attitudes, so long as one recognises one has no right to make such exceptions. Wallace's aim seems to be to describe a particular emotional dynamic. *Feeling* resentment against another who lied to one, when one is fully cognisant of having no qualms against lying to the other oneself, simply *is* an expression of a sense of

superiority, so long as it is not tempered with guilt. And hypocritical criticism is justifiably received with indignation by someone who is aware of her critic's hypocrisy.

But what of the case of *secret* hypocrisy, that *is* accompanied by a private recognition of one's own inconsistency, and an internal vow to avoid the morally troubling conduct in question from then on? Does one do wrong simply by issuing a hypocritical moral address? Wallace's suggestion is that respect for the principle of equal consideration is not simply a matter of one's beliefs, it is a matter of what one does. If one exposes another to the opprobrium of blame, but does not expose oneself, for the same offense, then one is 'participating in a system of social sanction' in a way that distributes opprobrium unequally, whatever one believes about one's own blameworthiness compared to others.

But this conception of the dynamics of blame as a system for distributing the burden of opprobrium is problematic. It is not the case that there is some fixed quantity of opprobrium to go round, and that one fails to accept one's fair share when one makes a secret of one's wrongdoing (and if there were, then presumably it would be the concealment of wrongdoing, and not hypocrisy, that was unfair). The wrong, he wants to say, is not simply holding 'double standards', but 'apply[ing] double standards...by accepting a threshold for subjecting others to opprobrium that is lower than the threshold [one] appl[ies] in [one's] own case' (Ibid. 333 footnote).

But again, this distinction between holding a standard and applying it is difficult to construe. If one blames someone for failing to meet some moral standard, but refuses to expose oneself to blame by publicising one's own failure to meet that same standard, it is not clear that one has applied to oneself a different 'threshold' for being subjected to opprobrium. That would be to say that one would need to be guilty of a more serious fault if one were to blame oneself to the same degree as one blames the other. But this does not accurately describe our psychology: indeed, it may be that the worse one's offense, the less likely one

is publicly to acknowledge it. It is simply psychologically demanding to expect someone publicly to admit fault. It is perhaps equally demanding to expect someone to eschew participation in the economy of blame until she has done so – to make herself a pariah from normal expectations of interpersonal respect. Thus, hypocritical moral criticism that is not accompanied by a feeling of indignation that is recalcitrant to reflection may not be indicative of a failure to respect a principle of equal moral concern, and thus may not be wrong.

This small qualification to Wallace's otherwise quite convincing picture is important, for two reasons. For one, it makes room for the permissibility of important cases of moral criticism. When one country issues moral criticism against a hostile country, it is typical for the hostile country to respond by pointing to similar failings on the side of their adversary. In the Soviet era, the response "And you are lynching blacks!" entered the public consciousness as the standard retort to all actual and potential criticism levelled against the USSR by America, so much so that it became the punchline of jokes in the Eastern Bloc.[47] A casual glance at the Twitter feed of the Russian Embassy to the UK will be enough to show that this tactic is very much alive and well – the author highlights, for example, the UK's territorial dispute with Argentina in response to criticism about Russia's annexation of Crimea,[48] or US interference in the internal affairs of other countries in response to criticism of Russia's attempts to influence the 2016 Presidential election in the USA.[49] Here, it is natural to treat the interaction as a prime example of Shklar's systematic hypocrisy – hypocrisy that serves an ideological function, but is not morally problematic in

[47] Such an example is offered in passing in an essay by Vaclav Havel, 'A: Your subway does not operate according to the timetable, B: Well in your country, you lynch blacks'. He refers to it as a 'commonly canonized demagogical trick' (Havel 1980 10).
[48] Russian Embassy, UK (@RussianEmbassy). Twitter Post. 2nd April 2019, 6.44am. (https://twitter.com/RussianEmbassy/status/1113074785528020992, retrieved 18/04/19)
[49] Russian Embassy, UK (@RussianEmbassy). Twitter Post. 29th March 2019, 4.49am. (https://twitter.com/RussianEmbassy/status/1111596154972749824, retrieved 18/04/19)

itself. Yet it is clear that a form of moral address is at work here: those members of the international community who opposed Russia's annexation of Crimea accused Russia of wrongdoing. It is problematic to suppose that Russia's counter-accusation of hypocrisy really does undercut the standing of others to make a moral address. If we have to wait for representatives of a country that is without sin to cast the first stone, then it will prove near-impossible for international norms to be upheld. International norms, if they are to be meaningful, must be considered a free-standing source of authority, for precisely the reason that states will rarely be in a position to accept the moral authority of their adversaries. The same may be true of interpersonal morality: Wallace is right that the danger of hypocrisy should serve as a catalyst for self-reflection, but a moral community in which hypocrites lack standing to make moral addresses may in the end be one which is very poorly policed.

The key reason, though, is that Wallace's picture cannot therefore make sense of the apparent wrongness involved in exempting oneself from a general policy of one's own making. Whether certain forms of moral address are problematic would seem to depend on whether such address constitutes an unjust exercise of power. Dacher Keltner is among psychologists who have argued that people who have greater social power (as indicated, for example, by wealth) are more prone to unethical behaviour and rule-breaking than those with less power. One study measured how often vehicles stopped at a pedestrian crossing, where vehicle type was used as a proxy for social status, with models of car being ranked 1-5. All cars in the lowest status category stopped for pedestrians, while 45% of vehicle in the highest category failed to stop, and there was a positive correlation throughout the range. In another study, there was a clear correlation between participants' perception of their social class and the number of sweets participants were willing to take from a jar, when told the sweets were for children participating in another study, but that they could take some if they wanted (Piff et al. 2012, also Keltner 2017). All this points experimentally to the

conclusion Wallace proposed intuitively: that a tendency to make an exception of oneself in moral matters goes hand in hand with an inflated estimation of one's own value relative to others. But an inflated estimation of one's self-worth is not a moral wrong: it constitutes, perhaps, a vicious lack of humility, but it does not constitute wrongdoing until it causes one to break rules or to treat others unjustly. And as has just been argued, hypocritical moral address does not obviously constitute unfair treatment in itself.

The Hypocritical Exercise of Power: A Novel Account of Wrongful Hypocrisy

According to the tradition of civic republicanism, the exercise of "arbitrary" power is considered a wrongful infringement upon individual freedom. What constitutes an exercise of power, and when is such exercise arbitrary? Keltner et al. (2003) define power as an individual's relative capacity to modify others' states by providing or withholding resources or administering punishments. This is a definition of relative power, which is not directly amenable to our purposes, but it does help to demonstrate the connection between the exercise of political power and moral address: just as political power can 'modify others' states' through coercion, so too can moral address, through the imposition of social 'punishment' (this is clearly very close to what Wallace had in mind when characterising opprobrium as a 'system of social sanction'). Pettit defines the morally problematic exercise of power – which he calls 'domination', as i) the capacity to interfere, ii) on an arbitrary basis, iii) in certain choices that the other is in a position to make (Pettit 1999 52). 'Interference' is defined as intentionally 'mak[ing] things worse for [another]', including, importantly, 'manipulation', 'agenda-fixing' and 'the deceptive or non-rational shaping of peoples' beliefs and desires' (Ibid.). Interference is arbitrary when it is 'chosen or not chosen at the agent's pleasure' (Ibid. 55). In particular, on Pettit's view, interference is arbitrary when it is chosen or rejected without reference to the interests or preferences of the affected parties – which, he later clarified, requires democratic collective control (see

Pettit 2012). But this specific conception of the idea of freedom as non-domination can be viewed as part of a more general conception, classically associated with Aristotle, Livy and Harrington and articulated in Harrington's conception of an 'Empire of Laws, not of Men' (Harrington 1737 38). It is a feature of Kant's definition of a Republican constitution that it be established 'by principles of the dependence of all upon a single common legislation (as subjects)' (Kant 1991 99).

The contention, then, is that certain forms of hypocritical speech constitute exercises of power or interference, which are arbitrary and are therefore prima-facie wrong, insofar as they curtail freedom. It is further argued that these forms of interference sometimes meet a threshold level such that they should be judged to be wrong on balance. Hypocritical moral address is more likely to meet that standard in a political context, or when the addresser's power is relatively much greater than the addressee's. Let us return to our example of the politician who exempts her own children from the very education system that she is keen to promote. There are several hypocritical exercises of power going on simultaneously in this case. For one, there is power at a legislative level: banning private education while finding a work-around for herself. For another, there is power at the level of shaping public opinion: the politician influences the public by promoting the education system, changing their beliefs, or at least attempting to do so. And for another, there is power at the level of moral criticism: the politician implies (we may imagine) that those who send their children to private school are worthy of blame, effectively imposing social sanction on those that do so. By secretly sending her own children to private school, the politician makes these exercises of power *arbitrary,* in that they are not constrained by the universal application of some external public standard. By exempting herself from the effects of this exercise of power, she undermines the authority that would legitimate it.

This conception of wrongful hypocrisy therefore has the advantage of being better able to explain the wrongness of hypocritical moral criticism, in cases that were problematic for Wallace because they did not seem to involve any anti-egalitarian emotional response. By morally criticising or blaming another, one clearly interferes with her in Pettit's sense. Blame takes away options: by inviting opprobrium upon someone, one signals that it is appropriate to treat her differently, to withhold from her the assumption that she is a well-intentioned potential co-operator (Scanlon 2010 139-52). And because one exempts oneself from that same interference, one uses power arbitrarily - 'at [one's] pleasure' - and not by some exogenous norm. This conception also helps to shed light on Runciman's second-order hypocrisy – it helps to explain why the cases Runciman cites seem to be more than cases of mere deception or fraud. Manipulation is an exercise of arbitrary power, as through it, one takes advantage of a public standard that does not in fact apply. Cameron can be viewed as exercising power arbitrarily, if he influenced peoples' beliefs about his intentions, and their desire to support him, through means that merely suited him, but - because based on false pretences – both breached social conventions and were in principle incompatible with collective control. Swapping masks in politics strikes us as wrong, because if one appeals to one external standard – say, green political values - to gain influence with one section of society, and another standard - say, libertarian political values - to gain influence with another section, these standards cannot in principle be universally applicable because they are inconsistent with one another. This makes them arbitrary, because a rule that is not universally applied cannot be the product of public standards or collective control.

There is much to be clarified: we are exercising influence in all sorts of ways in our everyday interactions. Why do some of these interactions stand in need of justification? Pettit is clear that not all interference counts as domination, because not all interference is arbitrary.

Advising, cajoling, the setting of a particular example – all these are ways in which one can interfere with another, by changing her beliefs and desires, but in many cases, they are not arbitrary, simply because the addressee has a measure of control over their influence. The addressee can choose to disregard them. If the interlocutors are interacting on equal terms, the addressee can make counter arguments, she can demand clarification, she can question evidence. In a close possible world in which her interlocutor were trying to influence her beliefs and desires in a different direction, it is not the case that she would have been convinced, because the same means of persuasion would not have been available. The promotion of general policies that one does not wish to apply to oneself is an exercise of arbitrary power because it enables one who wields this power to impose policies at their pleasure. The thought is that the lawgiver would be unconstrained by the need to enact policy in the general interest, as she herself intends to evade the policy for personal gain. The promotion of policies one does not wish to apply *at all* is an exercise in arbitrary power because it permits the agent to influence people whatever their preferences happen to be, again, without having to be bound by considerations of the general good.

Frank Lovett (2016, 2010) disagrees with Pettit about the proper conception of 'arbitrariness' that best expresses the republican principle. Where for Pettit, non-arbitrariness implies collective control, for Lovett, an instance of interference is non-arbitrary simply to the extent that it is effectively constrained by some external authority (Lovett 2010 96). Here, the proposal is that it is not necessary to draw a sharp distinction between these two views, which are both expressions of the same important principle, realised within different background frameworks. In many important interpersonal cases, the two come together: disingenuously influencing another in a way that bypasses her agency is also an exercise of power that breaches public norms or standards.

An explanation is also owed as to why this form of morally problematic hypocrisy does not include the cases of the naughty teacher and the criminal judge, who, it was suggested earlier, should not be considered guilty of morally problematic hypocrisy. The reason is precisely that their behaviour *is* constrained by public rules or standards, which are not of their own making. Their role is simply to enforce the rules. Compare a case in which another such secretly criminal judge, instead of simply passing sentence in accordance with sentencing guidelines and precedent, takes the opportunity to lambast the defendant with moralistic invective. Here, we might be more suspicious as to whether the judge was simply performing her legally defined role. Rather, she is using her position to interfere with the defendant in a way that goes beyond what the law prescribes, exposing the defendant to additional opprobrium 'at her pleasure'. This explains why we are more inclined to regard such a judge as morally blameworthy.

Is the Climate Activist Guilty of Wrongful Hypocrisy?

Finally, we are in a position to consider whether environmental advocates need to be concerned about falling into this form of morally problematic hypocrisy. The key consideration here should be whether environmentalist speech or political action constitutes an unjustified exercise of *power*. In principle, a truly influential hypocrite, who had the capacity to cause others to change their behaviour, desires, or beliefs – say, about the need to reduce their individual emissions – but was not prepared to make the same changes in her own life, could indeed be considered morally blameworthy. In the final few minutes of *An Inconvenient Truth*, Gore made the following claim: 'each one of us is a cause of global warming, but each of us can make choices to change that with the things we buy, with the electricity we use, the cars we drive. We can make choices to bring our individual carbon emissions to zero. The solutions are in our hands. We just have to have the determination to make them happen' (Guggengheim 2006). There is a reading, and a

contextualisation of this statement according to which it would arguably constitute morally problematic hypocrisy according to the conception defended here, assuming reports of Gore's unusually high carbon footprint are accurate.[50] What is particularly noteworthy is that this segment was the only part of his presentation in which he addressed potential solutions to the problem of climate change. Given the strength of the foregoing exhortations regarding the gravity of the crisis, one might reasonably conclude that these recommendations for a potential solution were offered with similar urgency. In other words, it might reasonably be concluded that Gore's aim was to change, through his intervention, the beliefs, desires and behaviours of others with respect to the importance of reducing their individual carbon footprints to zero. Given his position of influence, and thus the potential effectiveness of this use of power, it could well be argued that his failure to modify his own behaviour in the same way renders that use of power arbitrary, given it indicates that it was not constrained by an external, universal standard.

Another example of genuinely wrongful environmentalist hypocrisy, it might be argued, is evident in the framing of the debate offered by organisations such as Population Matters. This organisation and others like it point to the correlation between population and total carbon emissions, and on that basis promote population control as their primary strategy for mitigation. The upshot, some commentators have argued, is that such organisations target countries with high birth rates for intervention, when these are often some of the world's poorest countries, generally with relatively very low per capita emissions.[51] This

[50] This reading takes much for granted and is offered more by way of example than by way of personal reproach. The section of the film can certainly be contextualised differently: it falls within a segment whose major theme seems to be the avoidance of defeatism and despair. Another reading would therefore be that Gore was simply attempting to show individuals that there was *something* they could do to help, rather than becoming fatalistic in the face of such a serious crisis. On this reading, then, Gore's intervention should not be considered an exercise of arbitrary power, as individuals were supposed to be able to take or leave the offer, at their discretion.

[51] See eg. George Monbiot 2009, (https://www.monbiot.com/2009/09/29/the-population-myth/, retrieved 18/04/19)

might rightly be regarded as shielding rich countries with low birth rates from criticism – countries in which these advocates of population control happen to be based - when the per capita emissions of rich countries may be orders of magnitude greater than those being targeted.[52] It is the potential power imbalance between influencers and those influenced in this discourse that raises a red flag, as a morally problematic exercise in hypocritical arbitrary power.

As well as degree of relative influence, when considering whether a case of hypocrisy meets a level of seriousness sufficient for it to constitute wrongdoing according to this standard, we should also consider the seriousness of the changes that the agent produces in the addressee of their hypocrisy, through their power. Or rather, these are two routes to the same conclusion: the more powerful an agent is, the greater their capacity to effect changes, and the greater the changes they make, the more powerful they are. Exercises of power may be formally arbitrary, but have so little substantive effect on the dominated party that we would not consider them to be of particular moral concern. If a hypocritical agent, through her example or her critical speech, is unable to make much impact upon others, we would hesitate to call her hypocrisy wrongful. If, for example, an average individual espouses the goal of reducing personal emissions and tuts at her friends and colleagues for failing to recycle or for eating meat, while continuing to eat meat and regularly failing to recycle herself, we would be unlikely to judge that her hypocrisy constitutes wrongful arbitrary interference. This can be explained by the fact that she and her associates are relatively equally positioned in terms of power, so that her influence is not disproportionately difficult to resist. Her disapprobation does not carry an especially great

[52] It might be remarked that this characterization of the case places the form of hypocrisy described in conceptual territory very near to the concept of 'moral schizophrenia' raised by Steven Gardiner (2011), after Michael Stocker (1976). In the spirit of Wallace, we may judge that this connection can be regarded as an advantage rather than a fault.

social cost; her example does not have particularly high status as a model for imitation. Thus, although given a certain specification of the case, such an agent might be blameworthy for the reasons of the kind Wallace adduces, we need not worry that the conception of wrongful hypocrisy defended here risks extending concerns about arbitrary power to trivial cases.

How worried should we be about Wallace's hypocrisy, in a case like the one just described? As we have seen, Wallace's concern is grounded in a particular emotional response: resentment that is unresponsive to reflection. Without a full description of the case as a dialogic interaction, therefore, we cannot say for certain whether what the above agent does is wrongful – if, when her hypocrisy is pointed out to her, she immediately admits her mistake, then she can be found guilty of nothing more than commonplace human inconsistency, and if she clarifies that she feels no resentment for those she chides, but merely wishes to give advice, then there are no grounds to regard her stance as wrongful on Wallace's account. Anti-hypocritical jibes from the likes of Kelly probably function by attempting to attribute Wallace-style debasing hypocrisy to the environmentalist, painting them as haughty and arrogant. But the number of environmentalists who feel genuine hypocritical resentment even on reflection is probably very small, so this line of attack is easily parried. The hypocrisy of arbitrary power would seem to be the next best explanation for the wrongfulness of hypocrisy in such cases, as this view makes sense of the intuition of unfairness alluded to by Wallace, without requiring the perpetrator to have any particular emotional response. But as we have just argued, it is highly unlikely that the judgement according to which we find such hypocrisy to be wrongful would apply to individual environmentalists.

Accounting this form of hypocrisy as morally wrongful does have a potentially unwelcome result: that norms of hypocrisy avoidance would in some cases conflict with norms derived

from consequentialist reasoning. For example, it was argued that whether Gore does wrong depends on whether he can be regarded as deliberately influencing people's beliefs, desires and behaviours. But if Gore does have such influence, and as a result, a significant number of people do attempt to reduce their emissions to zero, and make good progress with that project, then presumably, Gore on balance produces better consequences by delivering his hypocritical speech than by not delivering it. On a consequentialist account, therefore, we should judge it morally right that he deliver it. For those who want to maintain a common-sense pluralism about moral principles, we would have to find some means of reconciling this potential conflict. But it does not appear that this will present a very great challenge. We could view the conflict as a balance of competing claims: the claims of certain people not to have their freedom curtailed by being subject to arbitrary power, against the claims of another group (conceivably not disjoint from the first) not to have their lives made worse by climate change, to the extent that Gore's intervention might prevent. While this is by no means an algorithmic calculation, we can make some general statements that constrain the problem, giving us reason to suppose it is not in principle intractable. For example, we can argue that if Gore foresaw his speech would have a considerable impact on mitigation, there is a strong case that this should constitute an excusing condition on his use of arbitrary power. We are left with the somewhat paradoxical corollary that the more effective the use of power in causing people to change their behaviour – and thus the more serious the arbitrary interference - the greater the force of the consequence-based excuse. But at the very least, this potential conflict is not a specific feature of the account of hypocrisy offered here. It is rather a fundamental methodological dispute in moral philosophy that it is beyond the scope of this chapter to resolve.

Are we to conclude, then, that environmentalists should be much more concerned about the hypocrisy discourse stirred up by climate change-sceptical commentators than they

currently seem to be? Generally speaking, no. Judith Shklar's picture of hypocrisy and anti-hypocrisy as a discrete system of purely ideological conflict, detached from first-order moral practice, is still very much the most accurate way of mapping the topography of the vast majority of this discursive territory. Our moral concern about hypocrisy is focused upon those cases, such as they are, in which a hypocritical environmentalist holds the balance of power in the discourse in question. It is telling that climate sceptics are keen to present environmentalists as a global hegemony, conspiring to shut down dissent in order to protect their own interests – if this were really so, the case for the moral wrongness of environmentalist hypocrisy would be much stronger. In reality, though, cases of this kind account for a tiny fraction of those targeted by climate-sceptic anti-hypocritical discourse. There are a small number of political contexts in which environmental hypocrisy is a genuine concern; in such cases it is often already widely acknowledged that the hypocrisy in question constitutively involves unjust power relations, as in the case of Population Matters. In the majority of cases, however, the balance of power comes down very much in favour of the opponents of climate change mitigation policies. All this can be seen as counting against a view, along the lines of Hourdequin's, according to which environmentalists can be regarded as having an especially strong duty to reduce their personal emissions, grounded in the moral value of hypocrisy-avoidance. Such a norm may arise out of strategic concerns, in combination with a first-personal commitment to combatting climate change, or out of shared participatory intention or joint commitment. In most cases, however, the average environmentally minded person should not be concerned that they are guilty of wrongdoing as a specific result of their hypocrisy.

7. Conclusion

At the pre-industrial baseline of date of 1750, the global average concentration of carbon dioxide in the atmosphere is estimated to have been 227 ppm (Freidlingstein et al. 2019). In 1990, the highest monthly average concentration recorded at the NOAA observatory in Hawaii was 357.32 ppm. The average concentration recorded for April 2020 was 416.21 ppm.[53] This means that about 45% of the total increase in the atmospheric concentration of CO_2 since the pre-industrial period has occurred since 1990, the year that the Intergovernmental Panel on Climate Change delivered its first assessment report. As 350 ppm is the highest atmospheric concentration of CO_2 that is considered "safe", 1990 also marks roughly the year in which emissions could for the first time be determinately associated with harmful impacts (see Hansen et al. 2008). Thus, in this sense, all of the 'dangerous' contributions to GHG emissions have taken place at a time when climate change was already well understood.

For these reasons, it is difficult to shake the intuition that some agent or agents ought to bear at least some degree of outcome responsibility for at least some proportion of the harms of climate change. Human agents have caused climate change, and a large proportion of those impacts were foreseen. But countervailing considerations are also strong. Outcome responsibility is linked to considerations of the fair distribution of costs. In the societies in which we live today, the fossil fuel economy has extended its tendrils into all aspects of our lives, so much so that extricating ourselves from them would be an extremely costly prospect. Individuals who rely on fossil fuels for so many elements of their

[53] Datasets available from the Global Monitoring Laboratory, https://www.esrl.noaa.gov/gmd/dv/data/index.php?category=Greenhouse%2BGases (retrieved 8 Oct. 2020)

day to day wellbeing have a powerful case that it would be unfair if remedial responsibility for the mitigation of climate change were imposed upon them. This is the paradox which formed the object of our discussion

We started with a negative argument. Chapter 2 gave a schematic case as to why a number of standard approaches in moral theory were inadequate to the problem. Chapter 3 argued that some of the most influential arguments for individual direct duties to refrain from contributing to GHG emissions are unpersuasive. While it may be the case that an individual performing an action which directly produced GHG emissions would increase the risk of harm by some small amount, this is not enough to show that there is a duty not to emit, as this risk has to be weighed against competing considerations. The claim that individuals have a duty to refrain from contributing to emissions on the grounds that by doing so they would become a necessary member of a group that jointly triggers significant harm either threatened to leave the individual implausibly culpable for the whole of climate change, or collapsed into an expected harm view.

Because the case for individual direct duties to mitigate one's emissions appears weak, if duties are to be assigned to individuals, the most promising method of doing so is an indirect one, where duties are first assigned to a group, and then individual duties are read off from group duties. Because the group of emitters, taken as a whole, is an uncoordinated group, it is difficult to see how duties could coherently be assigned to it, and thus difficult to see why duties should be viewed as trickling down to individual group members. The solution was to invoke a partial form of group agency, one that was rich enough to ground group level and corresponding individual-level duties, but thin enough to accurately capture the relations which hold between individual members of polluter groups.

That form of agency was quasi-participatory intention. For accountability to descend from the group to the individual level, it is enough that individuals intend to participate in a

practice that is collective in a minimal sense, for example the practice whereby carbon-intensive industries are affirmed as good. By intentionally affirming and reproducing the norms of that practice, one is participating in the minimal collective act of continuing the practice. Thus, it is appropriate to regard oneself as accountable, in the backward-looking sense of outcome responsibility, for the climate impacts of the practice, just as one would be accountable for one's participation in a joint enterprise.

In Chapter 5 we illustrated this through the example of the practice of SUV driving. Let us take another example. The low-cost aviation boom in Europe, precipitated by the easing of regulatory barriers through the common European legal framework, has transformed the way millions of Europeans live their lives, especially in the UK and Ireland. Cheap oversees air travel is now a basic expectation for most UK citizens, so much so that the idea of going on holiday is virtually inseparable from the idea of travelling by air. It is undeniable that many, if not most of us have adopted an affirmative stance towards this industry, and thus reproduce the norms that allow it to flourish. We dream of sunny getaways to punctuate the working year. Our social media feeds are full of photographs of exotic destinations. If we are ever forced to list our hobbies for some biography or online profile, we invariably write "travel". Aviation is the fastest growing industry in the EU in terms of contribution to GHG emissions, and understandably so – air travel has become a kind of proxy for middle-class status, meaning that as peoples' economic standard of living has improved, they have soon become keen to join the ranks of air travellers. Yet this industry is producing serious damage: if the aviation industry were a country, it would rank among the top 10 emitters.[54] Because individuals reproduce the norms according to which cheap air travel is understood as an important aspect of our shared way of life, we should each

[54] See https://ec.europa.eu/clima/policies/transport/aviation_en, (retrieved 8th October 2020).

regard ourselves as bearing some degree of outcome responsibility for the negative impacts of that industry.

A number of questions remain to be answered. Our problem was framed as a kind of gap between the amount of anthropogenic harm that will be generated through climate change, and the proportion of that harm for which outcome responsibility, and with it, remedial duties, could reasonably be assigned. The account of quasi-participatory accountability we defended goes some way to closing this gap, but it does not go all the way. Some carbon emissions cannot be linked to practices which are affirmed by their participants. With respect to these emissions, Iris Marion Young is arguably correct in her assessment that remedial responsibility should be regarded as arising from considerations of collective ability and collective self-interest. Such duties will not be as stringent as duties arising from attributions of outcome responsibility, but this may be the best we can do.

The responsibility gap framing also raises problems of moral accountancy. In the Introduction, I suggested that outcome responsibility could be attributed to carbon majors on the grounds that they extracted and sold fossil fuels which, when consumed, in combination produced 67% of global culmulative industrial emissions. I also argued that individuals bear outcome responsibility for impacts arising from practices they affirm. These outcomes overlap. Thus, it might be argued a kind of problematic double-counting looms. Important work remains to be done to elucidate the conditions under which it is appropriate to assign outcome responsibility for extracting fossil fuels, and when it is appropriate to assign it for consuming fossil fuels. We can at least note, though, that the form of quasi-participatory accountability defended here need not be exclusive: the fact that responsibilities overlap is in no way incoherent. In potential situations of "overshoot", where the combination of actions carried out in fulfilment of remedial duties went beyond

what was strictly required, we might simply judge that considerations of fairness merited some evenly distributed "rebate" of the excess.

Finally, there might be a lingering worry that we've changed the subject in the course of our discussion, from the question of when responsibility can justifiably be attributed for a given outcome, to the question of the conditions under which individuals should acknowledge their own accountability, and be motivated by it. If so, this shift is constructive. The structural injustice framing of the problem captures an important insight, namely that most individuals are not in any straightforward sense causally responsible or liable for the kind of group-caused wrongs under consideration. But it also masks an important truth: assignments of backward-looking accountability – which, as we have seen, need not be exclusive – are perhaps the core structuring concepts of our moral life. They are immediately comprehensible, and are deeply motivating in a way that considerations of forward-looking collective responsibility may never be. This is why we have argued that moral revisionism should be resisted: in trying to fix what is not broken, we overcomplicate the problem, and cast it as more intractable than it need be. Climate change is challenging enough without adding the need to inculcate entirely novel moral concepts to our worries.

[Blank Page]

Bibliography

Adam, Smith. *An Enquiry into the Nature and Causes of the Wealth of Nations*. London: Strahan and Cadell, 1776.

Anderson, Kevin. 'The Inconvenient Truth of Carbon Offsets'. *Nature News* 484, no. 7392 (5 April 2012): 7.

Archer, David. *Long Thaw: How Humans Are Changing the Next 100,000 Years of Earth's Climate*. Princeton, N.J: Princeton University Press, 2010.

Arendt, Hannah. 'Collective Responsibility'. In *Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt*, edited by James William Bernauer. Distributors for the U.S. And Canada Kluwer Academic Publishers, 1987.

Attari, Shahzeen Z., David H. Krantz, and Elke U. Weber. 'Climate Change Communicators' Carbon Footprints Affect Their Audience's Policy Support'. *Climatic Change* 154, no. 3 (1 June 2019): 529–45.

Baatz, Christian. 'Climate Change and Individual Duties to Reduce GHG Emissions'. *Ethics, Policy and Environment* 17, no. 1 (2014): 1–19.

Barry, Christian, and David Wiens. 'Benefiting From Wrongdoing and Sustaining Wrongful Harm'. *Journal of Moral Philosophy* 13, no. 5 (2016): 530–552.

Beitz, Charles. *Political Theory and International Relations*. Princeton: Princeton University Press, 1979.

Bell, Derek, Joanne Swaffield, and Wouter Peeters. 'Climate Ethics with an Ethnographic Sensibility'. *Journal of Agricultural and Environmental Ethics* 32, no. 4 (2019): 611–632.

Borchers Arriagada, Nicolas, Andrew J. Palmer, David MJS Bowman, Geoffrey G. Morgan, Bin B. Jalaludin, and Fay H. Johnston. 'Unprecedented Smoke-related Health Burden Associated with the 2019–20 Bushfires in Eastern Australia'. *The Medical Journal of Australia* 213, no. 6 (23 March 2020): 282–83.

Bratman, Michael E. *Shared Agency: A Planning Theory of Acting Together*. Oxford University Press, 2014.

———. 'Shared Intention'. *Ethics* 104, no. 1 (1993): 97–113.

Broome, John. 'A Reply To My Critics'. *Midwest Studies In Philosophy* 40, no. 1 (1 September 2016): 158–71.

———. 'Against Denialism'. *The Monist* 102, no. 1 (1 January 2019): 110–29.

———. *Climate Matters: Ethics in a Warming World*. 1st Edition. edition. New York: W. W. Norton & Company, 2012.

———. *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell, 1991.

Buchanan, Allen, and Russell Powell. 'De-Moralization as Emancipation: Liberty, Progress, and the Evolution of Invalid Moral Norms'. *Social Philosophy and Policy* 34, no. 2 (2017): 108–135.

Budolfson, Mark Bryant. 'The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response'. *Philosophical Studies* 176, no. 7 (1 July 2019): 1711–24.

Butt, Daniel. 'On Benefiting From Injustice'. *Canadian Journal of Philosophy* 37, no. 1 (2007): 129–152.

Collins, Stephanie. 'Filling Collective Duty Gaps'. *Journal of Philosophy* 114, no. 11 (2017): 573–591.

Coon, Charli. 'Why President Bush Is Right to Abandon the Kyoto Protocol'. *The Heritage Foundation*. Accessed 8 October 2020. https://www.heritage.org/environment/report/why-president-bush-right-abandon-the-kyoto-protocol.

Costello, Anthony, Mustafa Abbas, Adriana Allen, Sarah Ball, Sarah Bell, Richard Bellamy, Sharon Friel, et al. 'Managing the Health Effects of Climate Change: Lancet and University College London Institute for Global Health Commission'. *The Lancet* 373, no. 9676 (16 May 2009): 1693–1733.

Crawford, James R. *The Creation of States in International Law*. Oxford: Oxford University Press, 2006.

Cripps, Elizabeth. *Climate Change and the Moral Agent: Individual Duties in an Interdependent World*. Oxford, New York: Oxford University Press, 2013.

———. 'Climate Change, Collective Harm and Legitimate Coercion'. *Critical Review of International Social and Political Philosophy* 14, no. 2 (1 March 2011): 171–93.

———. 'Collectivities without Intention'. Journal of Social Philosophy 42, no. 1 (2011): 1–20.

———. 'On Climate Matters: Offsetting, Population, and Justice'. Midwest Studies In *Philosophy* 40, no. 1 (2016): 114–28.

Dalton, Peter. 'Extended Action'. *Philosophia* 24, no. 3–4 (1995): 253–270.

Dan-Cohen, Meir. 'Responsibility and the Boundaries of the Self'. *Harvard Law Review* 105 (1 `January 1991): 959.

Darwall, Stephen L. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press, 2006.

Davidson, Donald. *Actions, Reasons, and Causes*. Oxford University Press, 2001a.

———. *Intending*. Oxford University Press, 2001b.

Downs, Anthony. *An Economic Theory of Democracy*. Harper, 1957.

Driver, Julia. *Consequentialism*. Routledge, 2011.

———. *Uneasy Virtue*. Cambridge University Press, 2001.

Eckersley, Robyn. 'Rethinking Leadership: Understanding the Roles of the US and China in the Negotiation of the Paris Agreement'. *European Journal of International Relations*, 11 June 2020, 1354066120927071.

Elson, Luke. 'Incommensurability as Vagueness: A Burden-Shifting Argument'. *Theoria* 83, no. 4 (2017): 341–363.

Elster, Jon. 'Rationality, Morality, and Collective Action'. *Ethics* 96, no. 1 (1985): 136–55.

Feinberg, Joel. 'Collective Responsibility'. *Journal of Philosophy* 65, no. 21 (1968): 674–688.

———. *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press, 1970.

Foot, Philippa. 'Does Moral Subjectivism Rest on a Mistake?' *Oxford Journal of Legal Studies* 15, no. 1 (1995): 1–14.

French, Peter A. *Collective and Corporate Responsibility*. *Collective and Corporate Responsibility*. Columbia University Press, 1984.

Friedlingstein, Pierre, Matthew W. Jones, Michael O'Sullivan, Robbie M. Andrew, Judith Hauck, Glen P. Peters, Wouter Peters, et al. 'Global Carbon Budget 2019'. *Earth System Science Data* 11, no. 4 (4 December 2019): 1783–1838.

Gallie, W. B. 'Essentially Contested Concepts'. *Proceedings of the Aristotelian Society* 56, no. 1 (1955): 167–198.

Gardiner, Stephen M. 'A Call for a Global Constitutional Convention Focused on Future Generations'. *Ethics & International Affairs* 28, no. 3 (2014): 299–315.

———. 'A Perfect Moral Storm: Climate Change, Intergenerational Ethics and the Problem of Moral Corruption'. *Environmental Values* 15, no. 3 (2006): 397–413.

———. *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*. Environmental Ethics and Science Policy Series. Oxford, New York: Oxford University Press, 2011.

———. 'Accepting Collective Responsibility for the Future'. *Journal of Practical Ethics* 5, no. 1 (2017): 22–52.

———. 'Climate Ethics in a Dark and Dangerous Time'. *Ethics* 127, no. 2 (21 December 2016): 430–65.

———. 'Geoengineering and Moral Schizophrenia'. In *Climate Change Geoengineering: Philosophical Perspectives, Legal Issues, and Governance Frameworks*, edited by Andrew L. Strauss and Wil C. G. Burns, 11–38. Cambridge: Cambridge University Press, 2013.

Gardiner, Stephen M., and Allen Thompson, eds. *The Oxford Handbook of Environmental Ethics*. Oxford Handbooks. Oxford, New York: Oxford University Press, 2016.

Gardiner, Stephen M., and David A. Weisbach. *Debating Climate Ethics*. Oxford University Press. Accessed 9 October 2020.

Gardner, John. 'III—Discrimination: The Good, the Bad, and the Wrongful'. *Proceedings of the Aristotelian Society* 118, no. 1 (2018): 55–81.

Giddens, Anthony. *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. Basingstoke: Palgrave, 1979.

Gilbert, Margaret. *Joint Commitment: How We Make the Social World*. Oxford University Press, 2013.

——. 'Shared Intention and Personal Intentions'. *Philosophical Studies* 144, no. 1 (May 2009): 167–87.

Glover, Jonathan, and M. J. Scott-Taggart. 'It Makes No Difference Whether or Not I Do It'. *Aristotelian Society Supplementary Volume* 49, no. 1 (1975): 171–209.

Goodin, Robert E. 'Selling Environmental Indulgences'. *Kyklos* 47, no. 4 (1994): 573–96.

Goodin, Robert E, and Avia Pasternak. 'Intending to Benefit from Wrongdoing'. *Politics, Philosophy & Economics* 15, no. 3 (1 August 2016): 280–97.

'Greenhouse Gas Removal'. The Royal Society; Royal Academy of Engineering, 2018. https://royalsociety.org/greenhouse-gas-removal.

Guggenheim, David, (Director). *An Inconvenient Truth.* Paramount Classics 2006.

Gunster, Shane, Darren Fleet, Matthew Paterson, and Paul Saurette. 'Climate Hypocrisies: A Comparative Study of News Discourse'. *Environmental Communication* 12, no. 6 (18 August 2018): 773–93.

Hameeteman, Elizabeth. 'Future Water (In)Security: Facts, Figures and Predictions'. *Global Water Institute*, 2013.

Hansen, J., M. Sato, P. Kharecha, D. Beerling, R. Berner, V. Masson-Delmotte, M. Pagani, M. Raymo, D. L. Royer, and J. C. Zachos. 'Target Atmospheric CO_2: Where Should Humanity Aim?' *The Open Atmospheric Science Journal* 2, no. 1 (31 October 2008): 217–31.

Hardin, Russell. *Collective Action*. Resources for the Future, 1982.

Hardoon, Deborah, Sophia Ayele, and Ricardo Fuentes-Nieva. 'An Economy for the 1%: How Privilege and Power in the Economy Drive Extreme Inequality and How This Can Be Stopped'. Oxfam, 2016. http://oxf.am/ZniS.

Harrington, James. *The Oceana and Other Works of James Harrington Esq; Collected, Methodiz'd and Review'd, with an Exact Account of His Life Prefix'd, by John Toland. To Which Is Added, An Appendix, Containing All the Political Tracts Wrote by This Author, Omitted in Mr. Toland's Edition*. London: A. Millar, 1737.

Haslanger, Sally. *Resisting Reality: Social Construction and Social Critique*. Oxford University Press, 2012.

Havel, Václav, and Michal Schonberg. 'On Dialectical Metaphysics'. *Modern Drama* 23, no. 1 (1980): 6–12.

Heede, Richard. 'Tracing Anthropogenic Carbon Dioxide and Methane Emissions to Fossil Fuel and Cement Producers, 1854–2010'. *Climatic Change* 122, no. 1 (1 January 2014): 229–41.

Heglar, Mary Annaise. "The big lie we are told about climate change is that it's our own fault". *Vox*. 27th Nov 2018. (https://www.vox.com/first-person/2018/10/11/17963772/climate-change-global-warming-natural-disasters, retrieved 18/04/19).

Held, Virginia. 'Can a Random Collection of Individuals Be Morally Responsible?' *Journal of Philosophy* 67, no. 14 (1970): 471–481.

Hill, Thomas E., Jr. 'Ideals of Human Excellence and Preserving Natural Environments'. *Environmental Ethics* 5, no. 3 (1983): 211–224.

Hiller, Avram. 'Climate Change and Individual Responsibility'. *The Monist* 94, no. 3 (2011): 349–368.

———. 'System Consequentialism'. In *Consequentialism and Environmental Ethics*, edited by Avram Hiller, Ramona Ilea, and Leonard Kahn. New York: Routledge, 2013.

Hobbes, Thomas. *The English Works of Thomas Hobbes of Malmesbury, Now First Collected and Edited by William Molesworth*. Vol. 3. London: J. Bohn, 1839.

———. *The English Works of Thomas Hobbes of Malmesbury, Now First Collected and Edited by William Molesworth*. Vol. 7. London: J. Bohn, 1839.

Honoré, Tony. *Responsibility and Fault*. Hart Publishing, 1999.

Hook, Leslie. 'Oil Majors Gear up for Wave of Climate Change Liability Lawsuits'. *The Financial Times*. 9 June 2019.

Hooker, Brad. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford University Press, 2000.

———. 'Rule Consequentialism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2016. https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule.

Hourdequin, Marion. 'Climate Change and Individual Responsibility: A Reply to Johnson'. *Environmental Values* 20, no. 2 (2011): 157–162.

———. 'Climate, Collective Action and Individual Ethical Obligations'. *Environmental Values* 19, no. 4 (2010): 443–464.

Hursthouse, Rosalind. 'Environmental Virtue Ethics'. In *Working Virtue: Virtue Ethics and Contemporary Moral Problems*, edited by Rebecca L. Walker and Philip J. Ivanhoe. Clarendon Press, 2007.

———. *On Virtue Ethics*. Oxford University Press, 1999.

Inman, Mason. 'Carbon Is Forever'. *Nature Climate Change*, 20 November 2008, 156–58.

IPCC. 'Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change', 2013.

Jackson, Frank. 'Group Morality'. In *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*, edited by J. J. C. Smart, Philip Pettit, Richard Sylvan, and Jean Norman. Blackwell, 1987.

Jamieson, D. 'Ethics, Public Policy, and Global Warming'. *Global Bioethics* 5, no. 1 (1992): 31–42.

———. *Reason in a Dark Time: Why the Struggle Against Climate Change Failed — and What It Means for Our Future*. Oxford, New York: Oxford University Press, 2014.

———. 'When Utilitarians Should Be Virtue Theorists'. *Utilitas* 19, no. 2 (2007): 160.

Johnson, Baylor. 'The Possibility of a Joint Communique: My Response to Hourdequin'. *Environmental Values* 20, no. 2 (2011): 147–156.

Johnson, Baylor L. 'Ethical Obligations in a Tragedy of the Commons'. *Environmental Values* 12, no. 3 (2003): 271–287.

Jubb, Robert. 'Contribution to Collective Harms and Responsibility'. *Ethical Perspectives* 19, no. 4 (1 December 2012): 733–64.

———. 'Participation in and Responsibility for State Injustices'. *Social Theory and Practice* 40, no. 1 (2014): 51–72.

———. 'Recover It From the Facts as We Know Them'. *Journal of Moral Philosophy* 13, no. 1 (2016): 77–99.

Kagan, Shelly. 'Do I Make a Difference?' *Philosophy & Public Affairs* 39, no. 2 (1 March 2011): 105–41.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1997.

———. *Kant: Critique of Practical Reason*. Translated by Mary Gregor. 2nd ed. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press, 2015.

———. *Kant: Political Writings*. 2 edition. Cambridge England ; New York: Cambridge University Press, 1991.

Kelly, Julie. 'The hypocrisy of climate change advocates'. *The Hill*. 1st June 2016. (https://thehill.com/blogs/pundits-blog/energy-environment/313090-the-hypocrisy-of-climate-change-advocates, retrieved 18/04/19).

Kelly, Thomas, and Sarah McGrath. 'Is Reflective Equilibrium Enough?' *Philosophical Perspectives* 24, no. 1 (2010): 325–359.

Keltner, Dacher. *The Power Paradox: How We Gain and Lose Influence*. Reprint edition. Penguin Books, 2017.

Keltner, Dacher, Deborah H Gruenfeld, and Cameron Anderson. 'Power, Approach, and Inhibition'. *Psychological Review* 110 (1 May 2003): 265–84.

Kingston, Ewan, and Walter Sinnott-Armstrong. 'What's Wrong with Joyguzzling?' *Ethical Theory and Moral Practice* 21, no. 1 (2018): 169–186.

Kleingeld, Pauline. 'Kant's Second Thoughts on Race'. *Philosophical Quarterly* 57, no. 229 (2007): 573–592.

Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge ; New York, NY, USA: Cambridge University Press, 1996.

———. 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations'. *Philosophical Perspectives* 6 (1992): 305–332.

Kutz, Christopher. *Complicity: Ethics Law Collect Age: Ethics and Law for a Collective Age*. Reissue edition. Cambridge: Cambridge University Press, 2007.

———. 'The Collective Work of Citizenship'. *Legal Theory* 8, no. 4 (December 2002): 471–94.

Kutz, Christopher L. 'Shared Responsibility for Climate Change: From Guilt to Taxes'. In *Distribution of Responsibilities in International Law*, edited by André Nollkaemper and Dov Jacobs, 341–65. Shared Responsibility in International Law. Cambridge: Cambridge University Press, 2015.

Lawford-Smith, Holly. 'Difference-Making and Individuals' Climate-Related Obligations'. In *Climate Justice in a Non-Ideal World*, edited by Clare Hayward and Dominic Roser, 64–82, 2016.

Lenman, James. 'Consequentialism and Cluelessness'. *Philosophy and Public Affairs* 29, no.4 (Autumn 2002): 342-370.

Lewis, David. *Convention: A Philosophical Study*. Oxford: John Wiley & Sons, 1969.

———. 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* 61 no.4 (1983): 343-377.

Lichtenberg, Judith. *Distant Strangers: Ethics, Psychology, and Global Poverty*. Cambridge, UK: Cambridge University Press, 2014.

———. 'Negative Duties, Positive Duties, and the "New Harms"'. *Ethics* 120, no. 3 (2010): 557–578.

List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford, New York: Oxford University Press, 2011.

Lomborg, Bjorn. *The Skeptical Environmentalist: Measuring the Real State of the World*. Reprint Edition. Cambridge ; New York: Cambridge University Press, 2001.

Lovejoy, Thomas E., and Carlos Nobre. 'Amazon Tipping Point'. *Science Advances* 4, no. 2 (2018): eaat2340.

Lovett, Frank. *A General Theory of Domination and Justice*. Oxford University Press, 2010.

———. 'Civic Republicanism and Social Justice'. *Political Theory* 44, no. 5 (1 October 2016): 687–96.

MacIntyre, Alasdair. *After Virtue: A Study in Moral Theory*. United States: University of Notre Dame Press, 1981.

MacKinnon, Catharine A. 'Not a Moral Issue'. *Yale Law & Policy Review* 2, no. 2 (1984): 321–45.

Markowitz, Ezra M., and Azim F. Shariff. 'Climate Change and Moral Judgement'. *Nature Climate Change* 2, no. 4 (April 2012): 243–47.

May, Robert M. 'Will a Large Complex System Be Stable?' *Nature* 238, no. 5364 (August 1972): 413–14.

McCusker, Kelly E., Kyle C. Armour, Cecilia M. Bitz, and David S. Battisti. 'Rapid and Extensive Warming Following Cessation of Solar Radiation Management'. *Environmental Research Letters* 9, no. 2 (January 2014): 024005.

———. 'Rapid and Extensive Warming Following Cessation of Solar Radiation Management'. *Environmental Research Letters* 9, no. 2 (January 2014): 024005.

McKeown, Maeve. 'Iris Marion Young's "Social Connection Model" of Responsibility: Clarifying the Meaning of Connection'. *Journal of Social Philosophy* 49, no. 3 (2018): 484–502.

Miller, David. *National Responsibility and Global Justice*. Reprint edition. Oxford University Press, USA, 2012.

———. 'Taking Up the Slack? Responsibility and Justice in Situations of Partial Compliance'. In *Responsibility and Distributive Justice*, edited by Carl Knight and Zofia Stemplowska, 230–45. Oxford University Press, 2011.

Murkowski, F.H. 'The Kyoto Protocol Is Not the Answer to Climate Change'. *Harvard Journal on Legislation* 37, no. 2 (1 June 2000): XII.

Nefsky, Julia. 'Collective Harm and the Inefficacy Problem'. *Philosophy Compass* 14, no. 4 (2019): e12587.

———. 'Consequentialism and the Problem of Collective Harm: A Reply to Kagan'. *Philosophy & Public Affairs* 39, no. 4 (1 September 2011): 364–95.

———. 'Consumer Choice and Collective Impact'. In *The Oxford Handbook of Food Ethics*, edited by Mark Budolfson, Tyler Doggett, and Anne Barnhill, 267–286. New York, USA: Oxford University Press, 2018.

———. 'Fairness, Participation, and the Real Problem of Collective Harm'. *Oxford Studies in Normative Ethics* 5 (2015): 245–271.

———. 'How You Can Help, Without Making a Difference'. *Philosophical Studies* 174, no. 11 (2017): 2743–2767.

———. 'The Morality of Collective Harm'. UC Berkeley, 2012. https://escholarship.org/uc/item/9s49q22t.

Norcross, Alastair. 'Puppies, Pigs, and People: Eating Meat and Marginal Cases'. *Philosophical Perspectives* 18, no. 1 (2004): 229–45.

Nordhaus, William. *A Question of Balance*. Yale University Press, 2008.

Oppenheimer, M., M. Campos, R. Warren, J. Birkmann, G. Luber, B. O'Neill, and K. Takahashi. 'Emergent Risks and Key Vulnerabilities'. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global AndSectoral Aspects. Contribution of*

Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 1039–99. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

Otto, Friederike E. L., Ragnhild B. Skeie, Jan S. Fuglestvedt, Terje Berntsen, and Myles R. Allen. 'Assigning Historic Responsibility for Extreme Weather Events'. *Nature Climate Change* 7, no. 11 (November 2017): 757–59.

Parfit, Derek. *On What Matters: Volume One*. Oxford University Press, 2011a.

———. *On What Matters: Volume Two*. Oxford University Press, 2011b.

———. *Reasons and Persons*. Oxford: Oxford University Press, USA, 1984.

Peeters, Wouter, Derek Bell, and Jo Swaffield. 'How New Are New Harms Really? Climate Change, Historical Reasoning and Social Change'. *Journal of Agricultural and Environmental Ethics* 32, no. 4 (2019): 505–526.

Pettit, Philip. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge ; New York: Cambridge University Press, 2012.

———. *Republicanism: A Theory of Freedom and Government*. Oxford University Press, 1999.

Piff, Paul K., Daniel M. Stancato, Stéphane Côté, Rodolfo Mendoza-Denton, and Dacher Keltner. 'Higher Social Class Predicts Increased Unethical Behavior'. *Proceedings of the National Academy of Sciences* 109, no. 11 (13 March 2012): 4086–91. https://doi.org/10.1073/pnas.1118373109.

Pinkert, Felix. 'What If I Cannot Make a Difference (and Know It)'. *Ethics* 125, no. 4 (2015): 971–998.

Pogge, Thomas. "The Categorical Imperative", in Paul Guyer (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Rowman & Littlefield, 1997.

Posner, Eric A., and Cass R. Sunstein. 'Climate Change Justice'. *Georgetown Law Journal* 96 (2008 2007): 1565.

Rawls, John. *A Theory of Justice*. Harvard University Press, 1971.

———. *Lectures on the History of Moral Philosophy*. Cambridge, Mass: Harvard University Press, 2000.

———. *Political Liberalism*. New York: Columbia University Press, 1993.

Raz, Joseph. *Ethics in the Public Domain*. Oxford University Press, 1995.

———. *The Morality of Freedom*. Oxford University Press, 1988.

Regan, Tom. *The Case for Animal Rights*. University of California Press, 1983.

Rigaud, Kanta Kumari, and World Bank Group. *Groundswell: Preparing for Internal Climate Migration*, 2018. https://openknowledge.worldbank.org/handle/10986/29461.

Runciman, David. *Political Hypocrisy: The Mask of Power, from Hobbes to Orwell and Beyond*. Princeton University Press, 2009.

Sandler, Ronald. 'Environmental Virtue Ethics: Value, Normativity, and Right Action'. In *The Oxford Handbook of Environmental Ethics*, edited by Stephen M. Gardiner and Allen Thompson. Oxford ; New York: Oxford University Press, 2016.

——. 'Ethical Theory and the Problem of Inconsequentialism: Why Environmental Ethicists Should Be Virtue-Oriented Ethicists'. *Journal of Agricultural and Environmental Ethics* 23, no. 1 (2010): 167.

Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning, Blame*. Reprint edition. Cambridge, Mass.: Belknap Press, 2010.

——. *What We Owe to Each Other*. New Ed edition. Cambridge, Mass.: Harvard University Press, 2000.

Scheffler, Samuel. 'Relationships and Responsibilities'. *Philosophy and Public Affairs* 26, no. 3 (1997): 189–209.

Sheehy, Paul. *The Reality of Social Groups*. Routledge, 2016.

Shklar, Judith. 'Let Us Not Be Hypocritical'. *Daedalus* 108, no. 3 (1979): 1–25.

Shue, Henry. 'Climate Dreaming: Negative Emissions, Risk Transfer, and Irreversibility'. *Journal of Human Rights and the Environment* 8, no. 2 (1 September 2017): 203–16.

——. 'Global Environment and International Inequality'. *International Affairs* 75, no. 3 (1 July 1999): 531–45.

——. 'Historical Responsibility, Harm Prohibition, and Preservation Requirement: Core Practical Convergence on Climate Change'. *Moral Philosophy and Politics* 2, no. 1 (2015): 7–31.

——. 'Subsistence Emissions and Luxury Emissions'. *Law and Policy* 15, no. 1 (1993): 39–59.

Sidgwick, Henry. *The Methods of Ethics*. Edited by Jonathan Bennett. www.earlymoderntexts.com, 2017.

Singer, Peter. *Animal Liberation*. Avon Books, 1977.

——. 'Ethics and Intuitions'. *Journal of Ethics* 9, no. 3–4 (2005): 331–352.

——. 'Famine, Affluence, and Morality'. *Philosophy and Public Affairs* 1, no. 3 (1972): 229–243.

——. *Practical Ethics*. Cambridge University Press, 2011.

——. 'Utilitarianism and Vegetarianism'. *Philosophy & Public Affairs* 9, no. 4 (1980): 325–37.

Sinnott-Armstrong, Walter. 'It's Not My Fault: Global Warming and Individual Moral Obligations'. In *Perspectives on Climate Change*, edited by Walter Sinnott-Armstrong and Richard Howarth, 221–253. Elsevier, 2005.

Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.

Spiekermann, Kai. 'Buying Low, Flying High: Carbon Offsets and Partial Compliance'. *Political Studies* 62, no. 4 (1 December 2014): 913–29.

Srinivasan, Amia. 'Genealogy, Epistemology and Worldmaking'. *Proceedings of the Aristotelian Society* 119, no. 2 (2019): 127–156.

Stilz, Anna. *Liberal Loyalty: Freedom, Obligation, and the State*. Princeton: Princeton University Press, 2009.

Stocker, Michael. 'The Schizophrenia of Modern Ethical Theories'. *Journal of Philosophy* 73, no. 14 (1976): 453–466.

Strawson, P.F. *Freedom and Resentment and Other Essays*. Abingdon: Routledge, 2008.

Tanzi, Attila. 'Liability for Lawful Acts'. In *Max Planck Encyclopedia of Public International Law (MPEPIL).*, edited by Rüdiger Wolfrum, 2013. www.mpepil.com.

Tuck, Richard. *Free Riding*. Cambridge, Mass: Harvard University Press, 2008.

UNESCO World Water Assessment Programme. *Water in a Changing World: The United Nations World Water Development Report 3*. Paris: UNESCO, 2009.

UNICEF. 'Thirsting for a Future: Water and Children in a Changing Climate'. UNICEF, 2017. https://www.unicef.org/publications/index_95074.html, retrieved 17/12/20.

Wallace, R. Jay. 'Hypocrisy, Moral Address, and the Equal Standing of Persons'. *Philosophy & Public Affairs* 38, no. 4 (2010): 307–41.

Wallace-Wells, David. "'The devastation of human life is in view': what a burning world tells us about climate change". *The Guardian*. 2nd Feb 2019. https://www.theguardian.com/environment/2019/feb/02/the-devastation-of-human-life-is-in-view-what-a-burning-world-tells-us-about-climate-change-global-warming, retrieved 18/04/19.

Wiggins, David. 'An Idea We Cannot Do Without: What Difference Will It Make (Eg. To Moral, Political and Environmental Philosophy) to Recognize and Put to Use a Substantial Conception of Need?' *Royal Institute of Philosophy Supplement* 57 (2005): 25–50.

———. *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford University Press, 1997.

Wilkinson, Richard, and Kate Pickett. *The Spirit Level: Why Equality Is Better for Everyone*. Penguin UK, 2009.

Williams, Bernard. 'Getting It Right'. *London Review of Books*, 23 November 1989. https://www.lrb.co.uk/the-paper/v11/n22/bernard-williams/getting-it-right.

———. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge University Press, 1981.

———. 'The Point of View of the Universe: Sidgwick and the Ambitions of Ethics'. In *The Sense of the Past: Essays in the History of Philosophy*, edited by Myles Burnyeat. Princeton University Press, 2006.

Wood, Allen. 'Humanity as an End in Itself'. In Derek Parfit *On What Matters: Volume 2.* New York: Oxford University Press, 2011.

——. 'Kant on Duties Regarding Nonrational Nature'. *Aristotelian Society Supplementary Volume* 72, no. 1 (1998): 189–210.

Yetter Chappell, Richard. 'There Is No Problem of Collective Harm', n.d.

Young, Iris Marion. 'Equality of Whom? Social Groups and Judgments of Injustice'. *Journal of Political Philosophy* 9, no. 1 (2001): 1–18.

——. 'Katrina: Too Much Blame, Not Enough Responsibility'. *Dissent Magazine*, 2006a.

——. 'Responsibility and Global Justice: A Social Connection Model'. *Social Philosophy and Policy* 23, no. 1 (2006b): 102–130.

——. 'Responsibility and Global Labor Justice'. *Journal of Political Philosophy* 12, no. 4 (2004): 365–388.

——. *Responsibility for Justice*. Oxford University Press USA, 2011.